University of Reading

Henley Business School

ICMA Centre

Essays on Economic Forecasting with Alternative Datasets

by

Evripidis Bantis

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Supervisors: Prof. Michael P. Clements
Prof. Andrew Urquhart

Reading, January 2024

Declaration of Original Authorship

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Evripidis Bantis

To my family

Acknowledgements

First and foremost, I am deeply grateful to my supervisors, Prof. Michael Clements and Prof. Andrew Urquhart. Their continuous support and mentorship have been invaluable throughout my doctoral journey, inspiring me to think creatively, refining my critical skills, and significantly boosting my professional career prospects. It goes without saying that without their guidance, this research would not have been possible.

I would like to express my appreciation to my Ph.D. examiners, Dr. Ivan Sangiorgi and Prof. Chris Brooks, for their time, expertise, and thoughtful feedback during the evaluation of this thesis.

A special acknowledgment goes to the ICMA Centre for their generous financial support, enabling access to essential resources crucial for my research. Also, I am thankful to the students who were part of my classes, contributing to a vibrant academic environment and enriching my learning experience.

I am sincerely grateful to my Ph.D. colleagues; their company and friendship have made this journey special not only within the university but also beyond, filled with numerous amazing experiences and genuine friendships.

To my parents, Athina and Sakis, and my sister, Areti, I am profoundly grateful for their constant support. Without them, I could not have become the person that I am today.

Last but certainly not least, I want to express my deepest appreciation to my partner, Elisavet. Words cannot capture the immense gratitude and warmth I feel for all her support since the very first day of my doctoral studies.

Abstract

This thesis contributes to the field of economic forecasting using alternative datasets. The first study explores the benefits of search data for nowcasting GDP growth in the U.S. and Brazil, focusing on the marginal contribution of Google Trends compared to macroeconomic predictors. We use a dynamic factor model to address the large number of predictors and the "ragged-edge" problem. Our findings reveal that factor models incorporating Google "categories" data provide advantages over traditional models, with similar benefits observed in both economies regardless of the variable-selection strategy in the factor model. Using more detailed Google Trends data beyond its predefined "categories" does not yield additional benefits.

In the second study, we assess the potential of internet search data to enhance forecasts of private consumption and its components. Commencing with an initial set of consumption-related keywords, we construct three Google Trends datasets, encompassing search queries semantically related to the original terms. Employing various models suitable for high-dimensional structures, we show that Google Trends effectively forecasts aggregate private consumption, especially over long-term horizons and for durable goods post-pandemic, with random forests proving the most effective.

The final chapter examines if alternative predictors from internet searches and news articles can improve forecasts of inflation uncertainty in the United States. We create a novel set of predictors using Google Trends and Bloomberg's News Trends with inflation-related keywords. Three measures of inflation uncertainty are derived, reflecting disagreements in price expectations among households, investors, and professional forecasters. Results indicate significant forecast improvements for households' uncertainty

over short horizons from Google and News Trends. However, macroeconomic predictors remain more valuable for investors' uncertainty, and neither data source effectively predicts professional forecasters' uncertainty. Most benefits of using Google and News Trends to forecast households' uncertainty have emerged recently, highlighting their importance during times of uncertainty.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the aftermath of the 2007-2009 Global Financial Crisis, researchers and policymakers have shown an increased interest in gathering data at a more granular level than in the past, both temporally and cross-sectionally. This interest stems from the recognition that traditional macroeconomic data sources, while robust, often fall short in capturing the evolving economic dynamics in a timely fashion. Therefore, the necessity to monitor rapid changes in the economic landscape in real-time as well as for making informed policy decisions has led economists to extensively explore a wide range of alternative data.

Building on this need for more nuanced and timely data, developments in computing power have played a crucial role. These technological advancements have revolutionised our ability to store, manipulate, and analyse the immense volume of data that the modern economy generates. Similarly, the widespread penetration of the internet into society has been instrumental in data collection. Every interaction online, be it a search on Google, a post on social media, or a purchase on an e-commerce platform, leaves behind a digital footprint. This evolution has resulted in the creation of vast and varied datasets.

The digital age has thus given rise to several key types of data. According to a taxonomy based on the data content provided by the statistics division of the United Nations Economic Commission for Europe (UNECE) there are three main types of emergent data: Social Networks, Traditional Business Systems, and the Internet of Things (IoT). Social Networks data encompasses a wide spectrum of digital interactions, from activities on social media platforms like Facebook and Twitter to blog posts, comments, and video content on sites like YouTube. It also includes data from search engine queries, shedding light on public interests and concerns. Traditional Business Systems data focuses on the digital records of business transactions, particularly those involving debit and credit cards, offering a detailed view into consumer behavior and spending patterns. Additionally, the Internet of Things represents a rapidly expanding frontier of data collection. In this space, data is extracted from machines and sensors integrated into various devices, providing real-time, granular insights into a myriad of activities and conditions (Kapetanios et al., 2018). This type of data offers a unique perspective on the interconnectedness of economic and social activities.

Drawing from the academic literature, it becomes clear that internet search data, with a particular emphasis on Google Trends, has been a prominent feature in macroeconomic research utilising non-traditional datasets. This platform, which tracks the frequency of specific search terms on Google's search engine, offers insights into consumer intentions and behaviors, particularly in the realm of consumer spending. The premise is that prior to making economic decisions, such as purchases, individuals often seek information online, thereby generating data that can be invaluable for nowcasting macroeconomic variables. Additionally, the frequency of internet searches, as reflected in Google Trends, can shed light on the broader economic dynamics. In times of uncertainty, there is a noticeable increase in information-seeking behavior as individuals strive to understand the economic conditions around them. Therefore, Google search data might contain valuable information that reflects not only the preparatory steps involved in the purchasing process and intentions but also the sentiment around consumers.

What specific advantages does Google search data offer over traditional macroeconomic indicators? Arguably, the most significant drawback of traditional macroeconomic indicators is their substantial publication delay, which poses challenges in obtaining timely economic insights. For instance, in the United States, official data on key economic indicators are typically released with a considerable delay. Specifically, the advanced estimate of GDP and its components are only available about a month after the reference quarter, while data on private consumption, are also released with a lag of approximately a month. This situation is further exacerbated in the case of most developing economies, where the publication of economic data often faces even greater delays, thereby amplifying the challenges in obtaining timely and accurate economic insights. In contrast, Google search data are readily available and are published with virtually no delay. This immediate availability starkly contrasts with the time lag inherent in traditional economic data releases, offering a more real-time perspective on economic activities.

Another significant advantage of Google Trends data is its high-frequency nature. Unlike traditional macroeconomic data, which are often released on a monthly or quarterly basis, Google search data are available on a daily frequency. This higher frequency allows for a more granular and up-to-date understanding of economic trends and consumer behavior. This aspect of Google Trends data is particularly valuable in rapidly changing economic environments, where traditional data sources may lag behind the current economic reality.

Expanding upon these aspects, the academic literature further emphasises the wide-ranging advantages of using Google Trends for forecasting purposes. According to Buono et al. (2017), these benefits include ease of data access and collection, efficient data management and processing, the likelihood of similar data being continuously available in the future, and the overall high quality of the data. Moreover, in developing economies, traditional economic indicators, such as opinion surveys, often face issues of unavailability, significant publication lags, or lower quality compared to those in developed economies. In such contexts, Google data emerges as a viable alternative for predicting consumer behavior. Its real-time publication and high data quality make it particularly useful. Furthermore, in regions where there is substantial internet usage and Google dominates the search engine market, Google Trends data can effectively serve as a proxy for consumer sentiment. This capacity to reflect consumer attitudes and intentions through search behavior is especially valuable for economic forecasting, offering insights that might not be captured by traditional economic indicators.

The integration of these diverse big data sources into macroeconomic forecasting models opens up new avenues for capturing the current state of economic variables accurately and promptly. However, the abundance of information within this alternative type of data poses considerable challenges to conventional econometric techniques. In the realm of macroeconomic time series, dealing with "Fat Data" has become a common challenge. This term is employed to describe situations characterised by a substantial cross-sectional dimension (big N) relative to a limited temporal dimension (small T). In such scenarios, traditional methods like least squares or maximum likelihood often yield inferior predictions due to high estimation uncertainty or become impractical when the number of variables surpasses the number of observations. Consequently, effectively leveraging a large cross-section of variables in macroeconomic forecasting models necessitates meticulous econometric treatments. In the existing literature, various econometric and statistical models have been extensively employed to effectively handle high-dimensionality in the predictor space. These models include, *inter alia*, dense (factor models), sparse (LASSO, elastic net), ensemble (bagging, CSR), and non-linear models (random forests) (Kotchoni et al., 2019; Medeiros et al., 2021). Each of these types of models inherently involves a reduction in dimensionality, either explicitly or implicitly, aimed at mitigating the risk of overfitting and maximising performance in out-of-sample forecasting.

The aforementioned developments lay the foundation for the current thesis. We contribute to the growing field of economic forecasting using alternative datasets, particularly Google Trends, and a range of statistical and machine learning techniques to predict several key macroeconomic indicators such as GDP, consumption, and inflation uncertainty. More specifically, in Chapter 2, published in the *International Journal of Forecasting*, we investigate the usefulness of Google Trends data in macroeconomic nowcasting, specifically its added predictive power over traditional data sources. The primary goal of this chapter lies in assessing the value of incorporating Google search series data for nowcasting GDP growth rates using dynamic factor models in the United States and Brazil. While current literature primarily focuses on developed economies, our work significantly extends this analysis to also include emerging market environments, thereby providing a more comprehensive understanding of the applicability and insights offered by Google

Trends data across diverse economic landscapes. In addition, in contrast to the existing literature which predominantly employs bridge equation models, our study utilises dynamic factor models. These models can be quite effective within a nowcasting exercise, as they read the real-time flow of data and adeptly handle the non-synchronous release of information. Also, our analysis includes assessing the effectiveness of various variable selection methods, such as the elastic net, least absolute shrinkage and selection operator (LASSO), and an adaptive version of LASSO. This approach aims to determine whether focusing factor models on targeted predictors can lead to improved forecast accuracy. Additionally, we adopt what we term the "most recent performance" (MRP) approach to construct sets of targeted predictors. This involves selecting the variable selection method that demonstrates the best local out-of-sample performance each time the balanced dataset of predictors is updated with new information.

Interesting findings are derived from our analysis. First, we observe that factor models effectively incorporate new information as it becomes available within the reference quarter, as the forecast errors exhibit a downward trend from forecasting to backcasting horizons. For the United States, dynamic factor models consistently outperform the AR(1) benchmark across all forecast horizons. In the case of Brazil, the models show superior performance primarily at nowcasting and backcasting horizons. Second, our analysis reveals that factor models utilising both economic indicators and Google Trends categories exhibit the strongest performance compared to the autoregressive model. Third, the effectiveness of variable selection methods is most pronounced at forecast horizons and tends to diminish as more information is incorporated throughout the reference quarter in nowcasting and backcasting horizons. Notably, the most substantial benefits from variable selection are observed when the models include both Google data and economic indicators, rather than being limited to only economic indicators. Fourth, we find that the main categories of Google Trends data are the most useful predictors of GDP growth rates. When subcategories are added to the model, the forecast gains diminish, suggesting that the information within subcategories is already encapsulated by the main categories. Finally, our results do not conclusively indicate that Google data are inherently more valuable for one country over the other. Instead, the utility of Google data in both Brazil and the United States appears to depend on the specific variable-selection method employed. This finding underscores the nuanced nature of using Google Trends data in economic forecasting and highlights the importance of the methodological approach in leveraging such data.

In Chapter 3, we attempt to shed light on the usefulness of several Google search data in forecasting private consumption and its components, namely durable goods, non-durable goods, and services. It expands upon existing studies by incorporating a wider array of Google Trends data, venturing into new types of data that have not been previously considered in the literature for forecasting private consumption. In addition, this chapter addresses some of the inherent limitations in Google Trends series and employs a diverse range of econometric and machine learning methods suitable for high-dimensional data.

Specifically, similarly to the first empirical chapter, we assess the contribution of search data against traditional macroeconomic predictors. Three distinct sets of search predictors have been utilised, namely, “related queries”, “categories”, and “keyword planner”. Empirical findings reveal interesting insights. First, when we compare models using only macroeconomic data against those incorporating Google Trends data, we observe some intriguing patterns. For durable goods expenditures, Google Trends data considerably improve forecast accuracy, particularly at the $h = \{3, 6, 9\}$ horizons, with improvements of up to 50% compared to macroeconomic predictors. In contrast, nondurable goods show similar performance between Google Trends and macroeconomic-based models, with some exceptions at the 9-month horizon, indicating that the benefits of Google search data in forecasting nondurables goods is rather limited. For services, occasional gains are seen at the 9-month horizon, while for aggregate consumption, significant improvements are mainly documented at the $h = \{6, 9\}$ horizons.

Second, the study demonstrates that models enforcing sparsity yield the highest forecast accuracy, in contrast to dense models like factor models, which underperform in out-of-sample testing. The random forest model consistently achieves the highest forecast accuracy across different Google Trends datasets, target variables, and horizons, topping over 40% of the forecast experiments. LASSO regression also shows strong performance, whereas complete subset regressions demonstrate weaker results. All models encounter significant accuracy challenges during the early stages of the Covid-19 pandemic, with marked increases in forecast errors coinciding with the initial consumption decline and subsequent economic recovery. Finally, the last part of the empirical analysis in Chapter 3 delves into the selection of Google Trends series throughout the entire out-of-sample period. This investigation reveals dynamic shifts in the variables included in the models, showing a temporal change in predictive factors. Before the pandemic, a wider array of variables was incorporated compared to the period post-2020, indicating an evolving landscape of predictive factors over time, possibly related to business cycle fluctuations.

In Chapter 4, we explore the predictive ability of alternative data sources in forecasting inflation uncertainty in the United States. The contribution of the final chapter lies in linking Google and News Trends datasets to forecast inflation uncertainty, introducing a novel approach in the literature and expanding the study into the realms of high-dimensional and machine learning models for forecasting inflation dynamics. Given the challenges in forecasting inflation uncertainty due to the absence of a definitive measure, we derive three uncertainty indexes, each reflecting the perspectives of different economic agents: households, professional forecasters, and investors. Our forecasting approach centers on two key alternative data sources: Google Trends and News Trends. Starting with “inflation” as a primitive keyword, we expand our dataset using Google’s “related-keywords” feature to include semantically related queries. To evaluate the effectiveness of these data sources, we conduct monthly out-of-sample forecasting exercises. We assess the accuracy of forecasts derived from Google and News Trends in comparison to traditional macroeconomic factors. Given the large number of predictors involved in some of our forecasts, we employ methodologies suited for data-rich environments, including both linear and non-linear models like bagging, complete subset regressions (CSR), and random forests.

Our empirical analysis yields several compelling insights. Initially, in forecasting households' inflation uncertainty, Google Trends predictors demonstrate notable accuracy at short-term horizons of one to three months, particularly with bagging and complete subset regression models. While these models significantly outperform those based on macroeconomic predictors in short-term forecasts, their precision decreases for longer horizons, aligning more closely with results obtained using macroeconomic factors. News Trends also show satisfactory performance, and in some instances, a combination of Google and News Trends achieves the highest forecast accuracy. However, in forecasting investors' inflation uncertainty, macroeconomic factors are consistently more accurate across all horizons except at the very short-term horizon of $h = 1$. For professional forecasters' inflation uncertainty, forecast errors are broadly comparable across alternative and traditional predictors, with macroeconomic factors excelling at shorter horizons, whereas Google Trends offer some advantages at longer horizons of $h = 9$ and $h = 12$.

Secondly, bagging models generally provide the most precise predictions, especially when the target variable relates to consumer surveys, where they, along with CSR models, yield the lowest forecast errors compared to the benchmark. Overall, bagging stands out for its superior forecasting performance. Thirdly, an analysis of predictability over time reveals that for forecasting households' uncertainty, macroeconomic factors were the primary informative source until approximately 2017-2019. However, during the Covid-19 pandemic and periods marked by increased uncertainty due to geopolitical tensions and supply shocks, Google and News Trends data become significantly more valuable. In contrast, when forecasting investors' uncertainty, macroeconomic factors maintain the highest accuracy throughout the out-of-sample period, especially for horizons ranging from three months to one year. Finally, to discern which Google Trends series are most predictive, we find that for consumer survey-based target variables, the most informative Google Trends series predominantly focus on queries about "inflation". In contrast, for uncertainty measures based on market indicators and professional forecasters, selected queries extend to additional terms concerning the "economy" and "interest rates".

Finally, in Chapter 5, we summarise the main results of the thesis, briefly outline the weaknesses, and provide suggestions for future research. We argue that overlooking data revisions could potentially distort empirical findings and that the examination of the predictive power of Google search data within a real out-of-sample forecast exercise might deepen our knowledge in this area. Additionally, while the thesis shows the benefits of integrating alternative data sources like Google Trends with various econometric and statistical models, it does not thoroughly analyse the underlying mechanisms driving these improvements. Therefore, improving the interpretability and the narrative behind those findings could be crucial in establishing the value of these data sources and methods in the field of economic forecasting. Lastly, as Chapter 4 discusses, the usage of disagreement-based series as a proxy for inflation uncertainty is a subject of debate in academic literature. In contrast, more direct measures of uncertainty, particularly those derived from survey data, tend to be more widely accepted, highlighting a promising avenue for future research.

Chapter 2

Forecasting GDP Growth Rates Using Google Trends in the United States and Brazil

2.1 Introduction

Macroeconomic nowcasting[1] has received much attention from policymakers and market practitioners who require an accurate reading of the state of the economy (recent, current, and prospective). Unlike financial variables collected at a higher frequency and published with little delay, key macroeconomic variables such as GDP are only available at lower frequencies such as quarterly, and are generally only published with a significant delay. For example, in the United States, the advance estimates of GDP and its components are only available a month after the reference quarter, and in some countries, the delays are longer. This means it may be possible to exploit higher–frequency indicators produced in a timely fashion to generate nowcasts and forecasts of macro–variables before the official estimates are released.

Traditionally, three sources of data have been considered for macroeconomic nowcast-

[1] *Now–casting* is defined as in Giannone et al. (2008) and involves prediction estimates of the present, the near future and the recent past. The term is a contraction of "now" and "forecasting" and originates in meteorology.

ing: (i) hard indicators, such as retail sales and industrial production, (ii) surveys of opinions and intentions, and (iii) high–frequency financial market data. However, in recent years, due to computer technology advancements and the advent of online information–gathering services, alternative data sources have become available, usually referred to as Big Data[2]. A popular source of Big Data for short–term macroeconomic forecasting is Google Trends, which provides information about the frequency with which a particular term is searched. Google search data may contain insights into consumer and other agents' plans and intentions, and perhaps especially consumer spending. Consumers may seek information on Google's search engine before making economic decisions regarding purchases, for example. Consequently, Google search data may constitute a valuable source of information for nowcasting macro variables.

The above developments set the scene for the current chapter. Our aim is to determine whether alternative high–dimensional datasets, such as Google search series, contain additional predictive power over and above that contained in traditional data sources, and our modelling and forecasting strategy is designed to address this question. It is worth noting that we are interested in the marginal additional benefit of Google search data because the analyst will typically have access to both sources of information. Due to the large number of explanatory variables that are under consideration, we use the dynamic factor model (DFM) framework of Giannone et al. (2008).

Several papers have employed Google search data to forecast specific macroeconomic variables such as private consumption (Vosen and Schmidt, 2011), unemployment and employment rates (Choi and Varian, 2009; D'Amuri and Marcucci, 2017; Borup and Schütte, 2022), and price levels (Seabold and Coppola, 2015). However, only a few papers examine the usefulness of Google data in forecasting the overall economic activity, such as Ferrara and Simoni (2019) and Götz and Knetsch (2019). In particular, Götz and

[2]The term "Big Data" was first used in the economics and econometrics literature in Diebold (2003) (for more information regarding the origins of this term see Diebold (2021)). IBM classifies Big Data into four categories ("4 Vs"): volume, variety, velocity, and veracity. Types of Big Data include Social Networks, Traditional Business Systems, and the Internet of Things (Kapetanios et al., 2018). Doornik and Hendry (2015) distinguish three shapes in numerical Big Data: "tall" datasets, where there are not so many variables, N, but many observations, T, with $T \gg N$, "fat" datasets, in which the number of variables exceeds the number of observations, $N \gg T$, and "huge" datasets, where there many variables and many observations, that is, extremely large N and T.

Knetsch (2019) find that Google search data can lead to more accurate GDP growth forecasts for the German economy, while Google data works better as an alternative to survey indicators rather than in addition to them. Ferrara and Simoni (2019) conclude that Google Trends are particularly valuable in nowcasting Eurozone's GDP during the first four weeks of the quarter since during that time there is a lack of information about variables related to the economy. Nevertheless, when official variables become available, the forecasting power of Google Trends data disappears. Overall, findings suggest that Google search data may constitute a fruitful set of information for nowcasting or short-term forecasting of macroeconomic variables.

However, a critical question we address is whether these findings for advanced economies, such as Germany and the Euro area, are replicated for emerging economies. There are opposing reasons suggesting that Google search data might be more or less valuable for less developed economies. In such economies, traditional data sources may be of a lower quality, or information may not be available or may be fragmented. This suggests Google Trends data may fill a void and be more valuable. Against this, the Google Trends data may themselves be less useful if lower rates of internet usage make the information less representative.

Our use of the dynamic factor model may improve upon the simple bridge equations employed in Götz and Knetsch (2019), and Ferrara and Simoni (2019). Although the choice between these models remains an empirical question, dynamic factor models may be more suited to nowcasting, since they can read the flow of data in real-time, and effectively cope with non–synchronous data releases ("ragged–edges")[3].

[3] The discussion in this part of the thesis references the earlier version of Ferrara and Simoni's work (Ferrara and Simoni, 2019), which was the most current and available resource during the primary research and writing phase of this chapter. We have since become aware that the cited paper has been revised (Ferrara and Simoni, 2023), including significant updates to both the methodology and findings. Specifically, the authors propose a two-step approach, called *Ridge after Model Selection*, for nowcasting macroeconomic variables using Google and traditional economic data, with their theoretical contribution highlighting the procedure's optimality properties. The first step involves preselecting Google variables conditional on official variables, specifically aiming at the macroeconomic aggregate targeted for nowcasting. Following this, the second step applies Ridge regularization to these preselected Google series and official variables, selecting the Ridge tuning parameter through Generalised Cross-Validation. The empirical section includes a nowcasting exercise for GDP growth rates applied to the United States, Euro area, and Germany. Findings show that Google data convey valuable information about the state of the economy, highlighting the benefits of combining these data with official statistics and underscoring the importance of pre-selection. However, it is observed that during recessionary periods, model

Factor models have become a workhorse model at central banks and other institutions for short–term forecasting due to this ability to deal with large – "ragged-edged" – datasets, and mixed–frequencies of monthly predictors and quarterly GDP rates. A parsimonious structure is achieved by summarising the information of the many data releases with a few common factors. Regarding the estimation method of the DFM, the two–step estimator of Doz et al. (2011) is employed, where in the first step model parameters are estimated by principal components using a standardised balanced dataset, and in the second step, the Kalman filter is used to update the estimates using an unbalanced dataset. There is a vast amount of studies that use factor models to forecast macroeconomic variables, *inter alia*, Stock and Watson (002b) and Giannone et al. (2008) for the United States, Schumacher (2010) for Germany, Barhoumi et al. (2010) for France, Schiavoni et al. (2021) for the Netherlands, Jansen et al. (2016) for Euro area, Panagiotelis et al. (2019) for Australia, Caruso (2018) for Mexico, Bragoli and Fosten (2018) for Brazil, Dahlhaus et al. (2017) for BRIC economies, and Luciani et al. (2018) for Indonesia. For further discussions regarding theoretical aspects, estimation techniques and empirical applications of dynamic factor models see Bai et al. (008b), Stock and Watson (2011), Barhoumi et al. (2014), and Doz and Fuleky (2020).

It is well established that including the largest available dataset in a forecasting context does not necessarily lead to more accurate predictions. As Boivin and Ng (2006) suggest, including more series to estimate common factors could be less beneficial for forecasting in cases where idiosyncratic components are cross–correlated. Furthermore, factors that are not "targeted" on the variable to be forecast may not perform well. Bai and Ng (008a) provide evidence that it may be sensible to pre–select variables from the dataset prior to calculating the factors. Kim and Swanson (2018) and Cepni et al. (2019) demonstrate the efficacy of dimension–reduction and shrinkage methods for forecasting for the U.S. and emerging economies.[4]

Informed by these findings, we use a number of variable selection and shrinkage meth-

specifications incorporating solely non-preselected Google data exhibit the highest nowcasting accuracy.

[4]For further discussions of dimension–reduction methods and forecasting see, *inter alia*, Schumacher (2010) and Bulligan et al. (2015).

ods to select variables that are relevant for nowcasting GDP growth rates. These are the elastic net, the least absolute shrinkage and selection operator (LASSO), and the adaptive LASSO. To avoid "look–ahead" bias[5], we do not apply the variable selection methods to the entire sample period. Instead, first, we extract targeted predictors by using only information available during the in–sample period (2005–2014), and second, we compute targeted predictors by employing a "most recent performance" (MRP) method, where every time a balanced dataset is updated with new information, the "best" variable selection method is selected according to local out–of–sample performance, i.e., the variable selection method which produces the lowest average RMSFEs in a series of one–step–ahead forecasts over the last four quarters.

It should be stressed that we focus on forecasting GDP growth rates for two reasons. First, GDP is typically available four weeks after the end of the quarter in developed economies (e.g., United Kingdom, United States), while in the case of emerging economies, it requires more than eight weeks (e.g., Brazil, India). Taking into account that GDP is usually used as the measure of activity in an economy and that it is closely followed by policymakers and market practitioners, timely and accurate estimations are of utmost importance.

We choose to study Brazil because it typifies the usual challenges with developing economies, in that data can be of lower quality than in developed economies, and soft indicators such as opinion surveys are often unavailable or published with significant lags. Hence, the Google series can be used as a supplementary source of data to measure consumers' interest in certain keywords since they are published in real–time and are considered of high quality. In addition, Brazil also contains some important features and consists of an interesting case to study the predictive ability of Google Trends data within an emerging market environment. Google is the dominant search engine based on market share, while the internet penetration rate is considered satisfactory, especially compared to the other developing economies.[6]

[5]We wish to avoid using information that would not be known at the time the forecast is made, as this would tend to exaggerate forecast performance. Throughout we use a real-time approach to respect the out-of-sample nature of forecasting.

[6]The most obvious choices for a leading emerging economy were China, India, Russia, and South

The results of our forecasting exercises provide a number of insights. First, as most of the literature suggests, dynamic factor models successfully incorporate new information as it becomes available, with forecast errors tending to decrease as we move from forecasting to nowcasting and backcasting. Second, for the U.S., estimates of factor models outperform a simple autoregressive benchmark at all horizons, while for Brazil, they primarily outperform the benchmark at nowcasting and backcasting. Third, factor models that utilise both economic indicators and Google Trends categories outperform by far the benchmark in both countries, establishing the value of "Big Data" in the form of Google Trends data for now(fore)casting GDP growth. Fourth, benefits from performing variable selection before the computation of common factors tend to arise mainly at one-quarter-ahead forecast horizons ($h = 1$), and their performance decreases as we incorporate more data. In addition, we observe more gains from pre–selecting predictors when we use both economic indicators and Google data rather than when we employ only economic indicators. The gains to including Google Trends data when the model already includes economic indicators are found to depend on the variable–selection strategy. Google search data provide gains when we construct factor models based on economic indicators and the main Google Trends categories – further disaggregation of the Google Trends data to consider the sub–categories is not helpful from a forecasting perspective. Finally, our findings do not strongly point in one direction in response to the question of whether Google data are more or less useful in an emerging economy such as Brazil (relative to the U.S.).

The remainder of this chapter is organised as follows: Section 2.2 discusses the methodological framework while Section 2.3 briefly describes the dataset. Section 2.4 presents the nowcasting design and empirical results. Section 2.5 summarises the main findings of this chapter. Finally, Appendix A provides additional details on the empirical exercise.

Africa. China and Russia were discarded since Google is not the dominant search engine in these countries, while India and South Africa were rejected because the internet penetration was deemed relatively low (below 50% in both countries, Statista 2018). On the other hand, Google is the dominant search engine in Brazil (97% market share, Statista 2018) and the internet penetration rate is around 70% (Statista, 2018).

2.2 Methodological Framework

This section analyzes the methodological framework employed in this paper. Section 2.2.1 presents the dynamic factor model as proposed by Giannone et al. (2008) while Section 2.2.2 presents the main benchmark model. Finally, Section 2.2.3 briefly discusses the variable selection methods used to construct "targeted" predictors before computing common factors.

2.2.1 Dynamic Factor Model

In this chapter, a dynamic factor model is employed to forecast real GDP growth rates as proposed by Giannone et al. (2008). Dynamic factor models summarise the information contained in the set of predictors using a few latent common factors. In particular, if we assume that n corresponds to the cross-sectional dimension of the dataset and T to the number of observations, the aim of a dynamic factor model is to separate each observation of a series, say X_t, into two orthogonal unobserved components. The first component, the *common component*, captures the cross–sectional comovements across series and is assumed to be a linear function of a few, r, latent common factors, with $r \ll n$. The second component, the *idiosyncratic component*, captures variable–specific features and is assumed to be serially and cross–sectionally uncorrelated with the common factors. To summarise, the dynamic factor model can be written as:

$$X_t = \Lambda f_t + \varepsilon_t \tag{2.1}$$

where $X_t = (X_{1t}, \ldots, X_{nt})'$ is an $(n \times 1)$ stationary process of n variables with $t = 1, 2, \ldots T$ observations, Λ is an $(n \times r)$ matrix of factor loadings, $f_t = (f_{1t}, \ldots, f_{rt})'$ is an $(r \times 1)$ stationary process of common factors, and $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{nt})'$ is an $(n \times 1)$ stationary process of idiosyncratic errors. The product Λf_t in Equation (2.1) denotes the common component of X_t. Common factors (f_t) and idiosyncratic component (ε_t) are considered to be orthogonal, that is, $E(f_t \varepsilon_s') = 0$ for any t and s. Factors can be modelled as a vector autoregressive (VAR) process of order p:

$$\begin{gathered} f_t = \Phi_1 f_{t-1} + \ldots + \Phi_p f_{t-p} + Bu_t, \\ u_t \sim WN\left(0, I_q\right) \end{gathered} \tag{2.2}$$

where $\Phi_1, \ldots, \Phi_p$ is an $(r \times r)$ matrix of autoregressive parameters, and B is an $(r \times q)$ matrix of full rank q, and u_t is a q dimensional white noise process of the shocks to factors. The idiosyncratic component is assumed to be orthogonal to common shocks. For more details see Doz et al. (2011).

The system of equations (2.1) and (2.2) can be cast in a state space representation, in which Equation (2.1) represents the *measurement equation* and describes the relationship between the observed predictor X_t and the unobserved common factor f_t, while Equation (2.2) denotes the *state equation* and explains how the unobserved factors are generated from their lags and innovations.

The dynamic factor model, as specified in Giannone et al. (2008), follows a two–step approach. Firstly, preliminary estimations of factor loadings $(\hat{\Lambda})$ and common factors $(\hat{f}_t)$ are derived by principal components, by using a standardised, balanced dataset. Then, the autoregressive coefficients $(\hat{\Phi}_j, j = 1, \ldots, p)$ are derived by using the estimated factors, $\hat{f}_t$ in a VAR(p) model. Secondly, the Kalman filter and Kalman smoother are employed to re–estimate the unobserved factors using the unbalanced dataset. In order to deal with the "ragged–edge" problem (i.e., missing observations at the end of the sample) the variance of the idiosyncratic component is set to infinity when X_t is not available. The factors are then projected to the future $(\hat{f}_{T+h|T})$ by estimating recursively a VAR model of order p as shown by Equation (2.2). Forecasts of the monthly factors are aggregated into quarterly frequency by employing the transformation from Mariano and Murasawa (2003) and therefore a forecast of GDP growth rate can be obtained by:

$$\hat{y}^Q_{T+h|T} = \hat{\beta}_0 + \hat{\beta}' \hat{f}^Q_{T+h|T} \tag{2.3}$$

where $\hat{f}^Q_{T+h|T}$ denotes the $r \times 1$ vector of quarterly factors.

2.2.2 Benchmark Model

As an additional way of evaluating the performance of the factor models, we use a simple autoregression as the benchmark:

$$y_t^Q = \mu + \sum_{i=1}^{p} \phi_i y_{t-i}^Q + u_t^Q \tag{2.4}$$

where y_t^Q denotes the quarterly growth rate of GDP, μ is a constant term, ϕ_i are the autoregressive parameters, u_t^Q is an error term, and the model order p is chosen so the error term is approximately white noise. As noted in Section 2.1, the key comparisons are between the different factor models, for example, between those which exploit Google Trends data, and those which do not. However, an AR model serves as a check on whether any information can improve on simply using lags of the variable itself. In nowcasting we would expect that information pertaining to the quarter we are forecasting would prove beneficial.

2.2.3 Variable Selection Methods

We use the dynamic factor model to capture the information in large sets of variables for forecasting/nowcasting. As shown by Bai and Ng (008a) and others, the forecasting performance of factor models can often be improved by a judicious choice of variables from which to extract factors, including choosing "targeted" predictors, i.e., choosing variables which are correlated with the target variable. Götz and Knetsch (2019) show the value of this approach for forecasting GDP with Google Trends data. Other studies which support the pre–selection of variables prior to factor estimation include Schumacher (2010), which find least angle regression with the elastic net is useful. Kim and Swanson (2018) also find in favour of "combination models", i.e., factor models combined with variable selection models.

We consider a number of dimension–reduction methods prior to the estimation of the factors. These are the Elastic Net (ENET), the least absolute shrinkage and selection operator (LASSO), and an adaptive version of LASSO (AdaLASSO). These techniques, which have been extensively used in the aforementioned papers, impose a penalty term and result in the coefficients on some putative explanatory variables being set to zero. Factors can then be estimated for the set of variables which attracted non–zero coefficients.

The least absolute shrinkage and selection operator (LASSO), introduced by Tibshirani (1996), is a regression method that performs simultaneously variable selection and regularization, and potentially could improve prediction accuracy and interpretation. The LASSO minimises the residual sum of squares by imposing a constraint on the sum of the absolute values of the coefficients to be less than a constant:

$$\begin{gathered}\hat{\beta}^{lasso} = \arg\min_{\beta} \sum_{t=1}^{T} \left(y_t - \beta_0 - \sum_{i=1}^{n} x_{ti}\beta_i \right)^2 \\ \textit{subject to } \|\beta\|_1 \leq \tau\end{gathered} \tag{2.5}$$

where y_t is the tth observation of the target variable, β_0 is an intercept, x_{ti} the tth observation of the ith predictor, β_i the corresponding coefficient, $\|\beta\|_1 \equiv \sum_{i=1}^{n} |\beta_i|$ denotes the $L1$ penalty, and τ represents the tuning parameter. An equivalent way to write the LASSO estimator in *Lagrangian form* is:

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \left\{ \frac{1}{2} \sum_{t=1}^{T} \left(y_t - \beta_0 - \sum_{i=1}^{n} x_{ti}\beta_i \right)^2 + \lambda \|\beta\|_1 \right\} \tag{2.6}$$

The *Lagrangian* multiplier λ, called the LASSO regularization parameter, determines the amount of shrinkage: when $\lambda = 0$ corresponds to the OLS estimator and when $\lambda \to \infty$ eliminates all coefficients. This modification indicates that some coefficients are set exactly to zero. This is a vital feature, especially when the set of predictors has a big data structure. Also, it should be noted that all variables have been standardised to avoid the estimate to rely on the units of measurement.

There are two ways to compute the value of λ: a time–series cross–validation scheme and by using information criteria, such as AIC or BIC. We compute the value of λ using the BIC adapted to LASSO, where the degrees of freedom are adjusted according to the methodology of Stein's unbiased risk estimation (Zou et al., 2007). In a simulation study, Smeekes and Wijler (2018), show that regularization parameters determined by BIC appear to lead more frequently to exact identification compared to cross–validation.

Zou (2006) provides evidence that under specific scenarios the LASSO variable selection could be inconsistent. Therefore, a modified version of the LASSO estimator has been introduced, called the *adaptive* LASSO, in which adaptive weights are used for penalizing different coefficients in the $L1$ penalty. In particular, adaptive LASSO uses a weighted penalty of the form $\sum_{i=1}^{n} w_i|\beta_i|$, where $w_i = 1/\left|\hat{\beta}_i\right|^{\nu}$, $\hat{\beta}_i$ is the ordinary least squares estimate, and $\nu > 0$ determines how much we desire to highlight the difference in the weights.

Although the LASSO estimator could be very useful in situations where there are many zero coefficients in the true model, there are some limitations that we need to consider. In cases where there is a significant correlation in the regressors, the ridge estimator outperforms the LASSO (Bai and Ng, 008a). Zou and Hastie (2005) identify two issues with respect to LASSO estimator. Firstly, when the cross-section dimension exceeds the number of observations, $n > T$, LASSO can choose only up to T variables, which could be a major drawback for a variable selection method. Secondly, when there are some predictors with high pairwise coefficients, LASSO chooses only one among these predictors and does not care which one is chosen. Thus, Zou and Hastie (2005) introduced the *elastic net* penalty:

$$\hat{\beta}^{enet} = \arg\min_{\beta} \left\{ \sum_{t=1}^{T} \left(y_t - \beta_0 - \sum_{i=1}^{n} x_{ti}\beta_i \right)^2 + \lambda_1 \sum_{i=1}^{n} |\beta_i| + \lambda_2 \sum_{i=1}^{n} \beta_i^2 \right\} \tag{2.7}$$

Similar to the LASSO estimator, the elastic net performs on the same time shrinkage and variable selection, but also it can choose groups of correlated predictors.

2.3 Data Description

This empirical chapter aims to produce forecasts, nowcasts, and backcasts of GDP growth rates for Brazil and the United States using large datasets of traditional economic indicators and Google Trends data. Figure A.1 shows the evolution of GDP growth rates. This section describes both types of data, the main categories, and analyses the advantages and issues that arise using Google search data.

2.3.1 Economic Indicators

The main dataset consists of 96 and 115 economic indicators for Brazil and the United States, respectively, and spreads over into ten groups: *Economic Activity, External Sector, Government Sector, Housing Market, Labor Market, Leading Indicators, Monetary Sector, Prices, Retail Sector, and Survey Indicators.* All economic indicators have been downloaded from Bloomberg's *Key Economic Indicators* category and cover the period from January 2005 to September 2019. Table 2.1 provides a brief summary about the number of selected economic indicators by category, while a complete list of variables alongside with description, publication delays and transformations applied to make them stationary is available in Table A.4 and Table A.5 in Appendix A.

Table 2.1: Summary of Economic Indicators by Category

	Brazil	United States
Economic Activity	11	13
External Sector	10	9
Government Sector	10	9
Housing Market	2	10
Labor Market	2	19
Leading Indicators	2	19
Monetary Sector	19	10
Prices	25	6
Retail Sector	13	6
Survey Indicators	2	14
Total	96	115

Variables from consumer and producer prices categories have been merged under the *Prices* category while the personal and monetary sectors have been merged under the *Monetary* category. Although financial variables could provide information in a timely manner, they have been discarded because their volatile nature may incorporate significant noise in our model, and also as Bańbura et al. (2013) show, financial indicators tend to have a limited role in nowcasting the overall economic activity when a large set of economic variables is included. It should be noted that all variables have been standardised by subtracting the mean and dividing by the standard deviation as a means to avoid overweighting of predictors with high variances when deriving the factors.

2.3.2 Google Trends

The aim of Google Trends is to provide data about the frequency that a particular keyword is searched over the total search volume, on a specific period, in a given geographical region. Google provides an index instead of the actual search volume numbers because of privacy reasons. The index awards a score that ranges between 0 and 100. A value of 0 implies that there was not enough data for this query, while a value of 100 indicates the peak popularity of the search term.

Google trends data are available without any publication delay and show the popularity of internet users' searches in real–time. The Google Trends data that we employ in this study spans from January 2005 to September 2019 on a monthly frequency. Google classifies search queries into 25 main categories (Table 2.2), and within each category, there is a further division of 272 subcategories in total. For instance, the query "Debt Management" would be allocated to the category of "Credit and Lending", which is a subcategory of Finance. As the Google data are not seasonally adjusted, we take the annual growth rates. Also, two abrupt breaks in the Google Trends series occur due to improvements in geographical assignments and data collection systems in January 2011 and January 2016. We adjust the data for these breaks by multiplying the post-break observations by the ratio of the local averages of the observations before and after the break.

From the literature emerges that there are plenty of advantages in using data like Google Trends for forecasting purposes, among others, more timely forecasts, easy way of data access and collection as well as the ease of data management and treatment, the high possibility that similar data will be available in the future, and the good quality in data (Buono et al., 2017). Furthermore, in developing economies, traditional economic indi-

Table 2.2: Google Trends: Main Categories

Arts & Entertainment	Autos & Vehicles	Beauty & Fitness
Books & Literature	Business & Industrial	Computer & Electronics
Finance	Food & Drink	Games
Health	Hobbies & Leisure	Home & Garden
Internet & Telecom	Jobs & Education	Law & Government
News	Online Communities	People & Society
Pets & Animals	Real Estate	Reference
Science	Shopping	Sports
Travel		

cators, such as opinion surveys, are sometimes unavailable or published with significant lags and are often considered to be of lower quality compared to developed economies. Thus, Google data can be used as an alternative source of data to predict consumers' behavior since they are published in real–time, and the data quality is high. Thus, in cases where there are high internet penetration rates and Google's search engine market share is dominant, Google data can be used as a proxy to consumer sentiment.

On the other hand, several issues have been identified when incorporating Google Trends data in a nowcasting framework. Firstly, according to Seabold and Coppola (2015), data from Google Trends are not the same over time, but instead, historical data from day to day can be different. This means that the sampling methodology that Google uses incorporates measurement error into the series. For a specific keyword on a particular day, Google provides precisely the same series, but for that specific keyword on different days, results are not the same (Carrière-Swallow and Labbé, 2013)[7]. Secondly, Google Trends series may exhibit strong seasonal components. Thirdly, there is a lack of knowledge of how Google treats and processes data. For example, queries are grouped in Google Trends categories by using a natural language algorithm, whose details are unknown (Kapetanios et al., 2018). Additionally, there are some drawbacks when Google

[7] To deal with this issue, D'Amuri and Marcucci (2017) take the simple average of Google Trends data for a specific keyword extracted from two different IP addresses and on 12 different days. Medeiros and Pires (2021) investigate the constantly changing Google Trends samples and highlight the importance of taking averages of several different samples to improve the consistency of the data series. In our setting, we rely on single downloads since we deal only with main categories and subcategories in which cross-correlations between different samples are always above 0.99.

Search data are used in emerging countries. According to Carrière-Swallow and Labbé (2013), there is no clear evidence of whether the internet has integrated into buyers decision process in developing economies. Moreover, a substantial proportion of household consumption consists of non–discretionary expenditure, where there is no need for a thorough search on Google engine before buying these products.

Taking into consideration that each query's meaning can change over time, it seems logical to incorporate the main Google search categories into our analysis. Additionally, we aim to shed light on the potential usefulness of disaggregate Google series, and therefore we also utilise the Google search subcategories, leading to a total of 297 Google series for each country.

2.4 Nowcasting Exercise

This section presents the nowcasting exercise and empirical results. Section 2.4.1 describes the nowcasting design, the different types of datasets, and the mechanics of variable selection methods. Section 2.4.2 exhibits the main empirical results, while Section 2.4.3 attempts to provide an answer regarding the potential significance of variable and model selection methods. Section 2.4.4 focuses on the forecast benefits that arise by including Google data in the factor model and finally, Section 2.4.5 presents the indicators that have been chosen by the variable selection model.

2.4.1 Setup

The forecasting performance of the factor model is assessed by a *pseudo* real–time out–of–sample exercise. We do take into consideration the publication delay of every variable in an attempt to avoid using data that would not have been available at the time of the forecast. However, due to the irregular publication pattern of the variables, we follow the approach of Giannone et al. (2008) and assume that publication delays are constant during the evaluation period. This assumption is not too unrealistic since variation in publication delays of the variables are only minor. Also, our approach does not take into account data revisions since real–time vintages for all the predictors in our dataset are not available. However, differences in data revisions are considered idiosyncratic, and therefore they do not affect the computation of common factors (Giannone et al., 2008).

For the United States, a sequence of seven predictions is produced for each quarter in the out–of–sample period, obtained in consecutive months. In particular, we generate three monthly one–quarter–ahead *forecasts* ($h = 1$), three monthly current quarter forecasts or *nowcasts* ($h = 0$), and one monthly *backcasts* ($h = -1$), i.e., forecasts for the preceding quarter. For Brazil, we generate a sequence of eight predictions, i.e., three monthly forecasts, three monthly nowcasts, and two monthly backcasts, since the publication delay of GDP for Brazil is around eight weeks, while for the United States is approximately four weeks.

Table 2.3 presents an example of timing predictions for Brazil for the first quarter of 2019. The first, the second, and the third forecasts are made at the end of October, November, and December of 2018. Thereafter, three nowcasts are estimated throughout the reference quarter at the end of January, February, and March 2019. Finally, we backcast the 2019:Q1 at the end of April and May, and in early June, the first official estimate for the GDP is being released.

GDP growth rate predictions are estimated recursively, where the first sample begins in January 2005 and ends in January 2014, and the last sample begins in January 2005 and ends in August 2019. Hence, the out–of–sample evaluation period is 2014:Q2 to 2019:Q2, i.e., 21 quarters. For the factor model specification, we optimise the number of factors and shocks at every step of our forecasting process, by employing information criteria from Bai and Ng (2002) and Bai and Ng (2007), respectively.

Let us assume that we have three types of data at our disposal: Economic Indicators ("EI"), Google Trends Categories ("GTC"), and Google Trends Subcategories ("GTS"), denoted by X_{EI}, X_{GTC} and X_{GTS}, respectively. Thus, as we have shown in Section 2.2.1 our aim is to forecast GDP growth rates, $y^Q_{T+h|T}$, based on direct factor forecasts, $f_{T+h|T}$. If we denote the available information set by Ω_T:

Table 2.3: Timing of Forecast Exercise for the First Quarter of 2019

	Forecast Type	Month	Forecast made on last day of
1	One-quarter ahead	1	October
2		2	November
3		3	December
4	Nowcast	1	January
5		2	February
6		3	March
7	Backcast	1	April
8		2	May

$$Proj[f_{T+h}|\Omega_T] \tag{2.8}$$

and the information set is defined as:

$$\Omega_T = \{X_{t,j},\ t = 1, \ldots, T, \text{and } j \in \{EI, GTC, GTS\}\} \tag{2.9}$$

The goal of this chapter is to evaluate whether Google Trends data can convey additional predictive power over and above that contained in traditional economic indicators. Thus, we first compute the common factors based only on economic indicators, and then we add Google Trends categories and subcategories to assess their marginal contribution. In particular, we first forecast common factors (f_{T+h}) based on economic indicators:

$$Proj[f_{T+h}|\Omega_T] \tag{2.10a}$$

$$\Omega_T = \{X_{t,j},\ t = 1, \ldots, T,\ j \in \{EI\}\} \tag{2.10b}$$

then we add Google Trends categories:

$$Proj[f_{T+h}|\Omega_T] \tag{2.11a}$$

$$\Omega_T = \{X_{t,j},\ t = 1, \ldots, T,\ j \in \{EI, GTC\}\} \tag{2.11b}$$

and finally, we add Google Trends subcategories:

$$Proj[f_{T+h}|\Omega_T] \tag{2.12a}$$

$$\Omega_T = \{X_{t,j},\ t = 1, \ldots, T,\ j \in \{EI, GTC, GTS\}\} \tag{2.12b}$$

Common factors are estimated based on datasets without variable pre–selection as well as based on targeted predictors. This chapter uses three variable selection methods to construct a set of targeted predictors: the Elastic Net, the LASSO, and the adaptive LASSO approach. To facilitate this, we follow two methods. First, targeted predictors are extracted using only information available in the in–sample period, i.e., 2005–2014. Second, we attempt to dynamically adjust the set of targeted predictors throughout the out–of–sample period by re–estimating the variable selection models each time we have new information and a balanced dataset. Additionally, our aim is not only to dynamically update the set of targeted predictors but also to select the "best" variable selection method according to their local out–of–sample performance; that is, we select the model which produces the lowest average RMSFE in a series of four one–step–ahead forecasts over the last four quarters. We call this approach the "Most Recent Performance" (MRP) method.

At each forecast origin, we forecast common factors based on observations of the predictors available at that period of time. For example, when our aim is to compute one–quarter ahead forecasts of 2019:Q1, we compute a series of three monthly forecasts in the last day of October, November, and December based only on information that is available on that dates, that is, 2005:M1 – 2018:M10, 2005:M1 – 2018:M11, 2005:M1 – 2018:M12, respectively. Specifically, we provide below a step–by–step analysis of our modelling framework in the case of the U.S. when the forecast origin is on the last day of October 2018 and when the MRP method is employed before estimating the DFM:

1. Undertake variable selection using X predictors for the period 2005:M1–2018:M7 (July 2018 is the latest available balanced dataset) and obtain X_s with $s = \{LASSO, AdaLASSO, ENET\}$

2. Conduct a series of four one–step–ahead out–of–sample forecasts for $y^Q_{T+1|T}$ where T=2017:Q2, 2017:Q3, 2017:Q4, 2018:Q1 using X_s with a DFM (steps of DFM are explained below). The set of predictors that produce the best–RMSFE is selected and denoted by X^*

3. Given the selected variables X^*_t, where t = 2005:M1–2018:M10, separate each observation into two orthogonal unobserved components:

$$X^*_t = \Lambda f_t + \varepsilon_t \tag{2.13}$$

4. Use PCA to obtain preliminary estimations of factor loadings ($\hat{\Lambda}$) and common factors ($\hat{f}_t$) using a standardised and balanced dataset ($\overline{X}^*_t$) where t=2005:M1–2018:M7

$$(\hat{f}_t, \hat{\Lambda}) = \arg\min_{f_t, \Lambda} (NT)^{-1} \sum_{t=1}^{T} (\overline{X}^*_t - \Lambda f_t)'(\overline{X}^*_t - \Lambda f_t) \tag{2.14}$$

5. Use Kalman filtering and smoothing techniques to re–estimate the factors for the unbalanced dataset (2005:M1–2018:M10). To deal with the ragged–edges, the variance of the idiosyncratic component is specified as follows:

$$E[\varepsilon^2_t] = \begin{cases} \psi, & \text{if } X^*_t \text{ is available} \\ \infty, & \text{if } X^*_t \text{ is not available} \end{cases} \tag{2.15}$$

and therefore no weight will be placed on missing observations in the estimation of common factors

6. Estimate the VAR on the period T=2005:M1–2018:M10, and forecast with the VAR one–step ahead, i.e. for 2018:M11:

$$\hat{f}_{T+1|T} = \hat{\Phi}_1 \hat{f}_T + \ldots + \hat{\Phi}_p \hat{f}_{T-p+1} \tag{2.16}$$

where T refers to 2018:M10. The two-step ahead forecast of 2018:M12 is given by:

$$\hat{f}_{T+2|T} = \hat{\Phi}_1 \hat{f}_{T+1|T} + \hat{\Phi}_2 \hat{f}_T + \ldots + \hat{\Phi}_p \hat{f}_{T-p+2} \tag{2.17}$$

for $p > 1$ and so on. That is, the forecasts are generated iteratively.

7. Forecasts of the factors at the quarterly frequency are calculated from the forecasts of the months using the Mariano and Murasawa (2003) transformation[8]

8. After the quarterly frequency of common factors is computed, bridge equations can be estimated (see for example Giannone et al. (2008)) since the target variable and the predictors have the same frequency, for the period 2005:Q1 to 2018:Q2 and use these to forecast the target variable for periods 2018:Q3 to 2019:Q1

$$y^Q_{T+h|T} = \hat{\beta}_0 + \hat{\beta}' \hat{f}^Q_{T+h|T} \tag{2.18}$$

9. Estimate AR benchmark using data from 2005:Q1 to 2018:Q2, and generate a three–step ahead forecast of the target variable for 2019:Q1:

$$Y_{T+3|T} = \hat{\alpha}_0 + \hat{\alpha} Y_T \tag{2.19}$$

Afterwards, we update again the model on the last day of November 2018 and perform exactly the same steps by incorporating the latest available data:

1. Undertake variable selection using X predictors for the period 2005:M1–2018:M8 (August 2018 is the latest available balanced dataset observed at the end of November 2018) and obtain X_s

2. Conduct a series of four one-step ahead out–of–sample forecasts for $y^Q_{T+1|T}$ where T=2017:Q3, 2017:Q4, 2018:Q1, 2018:Q2 using a DFM. The set of predictors that produce the best–RMSFE is selected and denoted by X^*

[8] Various other methods for handling mixed frequencies in forecasting are discussed in the literature. For further details see Foroni and Marcellino (2014).

3. Given the selected variables X_t^*, where t=2005:M1–2018:M11, separate each observation into two orthogonal unobserved components, i.e. the common and idiosyncratic component

4. Use PCA to obtain preliminary estimations of factor loadings and common factors using a standardised and balanced dataset $(\overline{X}_t^*)$ where t=2005:M1–2018:M8

5. Use Kalman filtering and smoothing techniques to re–estimate the factors using the unbalanced dataset (2005:M1–2018:M11) and deal with the ragged–edges

6. Estimate the VAR on the period 2005:M1–2018:M11, and forecast with the VAR for periods 2018:M12–2019:M3

7. Transform the monthly forecasts of the factors to quarterly using the Mariano and Murasawa (2003) transformation

8. Estimate bridge equations for the period 2005:Q1 to 2018:Q3 and use these to forecast the target variable for the period 2018:Q4 to 2019:Q1

9. Estimate AR benchmark using data from 2005:Q1 to 2018:Q3, and generate a two–step ahead forecast for the target variable for 2019:Q1

Finally, the information criteria from Bai and Ng (2002) are employed to determine the number of common factors:

$$IC(r) = lnV_r(\hat{f}_t, \hat{\Lambda}) + rg(N,T) \tag{2.20}$$

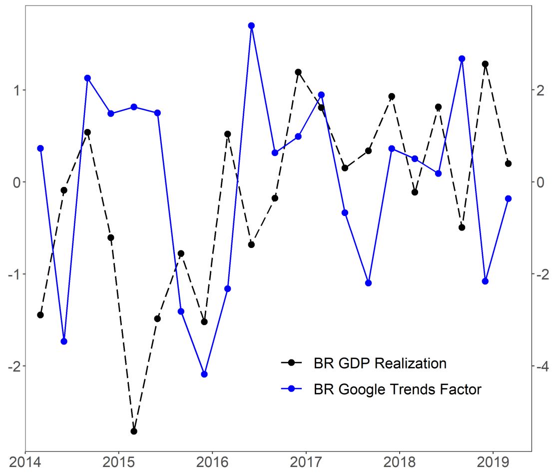
where $V_r(\hat{f}_t, \hat{\Lambda})$ represents the residual sum of squares objective function in Equation (2.14). In this function, r common factors are computed as the principal components, and $g(N,T)$ serves as the penalty term. While the IC_{p2} penalty function, introduced by Bai and Ng (2002) and defined as $g(N,T) = \left[\frac{N+T}{NT}\right] \ln\left[\min(N,T)\right]$, is typically employed, we also use the IC_{p1} criterion. This criterion is calculated as $g(N,T) = \left[\frac{N+T}{NT}\right] \ln\left[\frac{NT}{N+T}\right]$ and serves as a robustness check.

2.4.2 Forecasting Performance of Factor Models

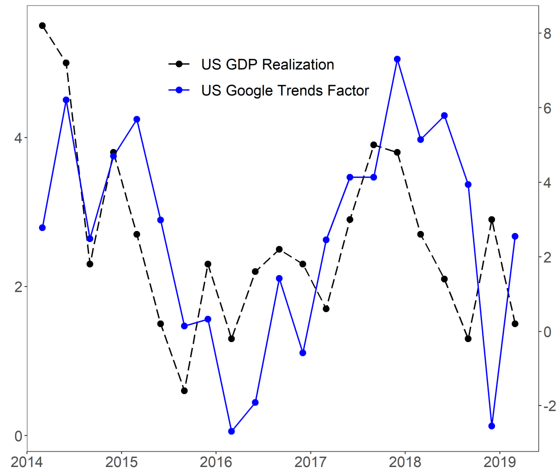
To assess the forecasting performance of the factor model, a *pseudo* real–time out–of–sample forecasting exercise is employed for the period of 2014:Q2 to 2019:Q2, and a sequence of eight (seven) predictions is computed for each quarter for the case of Brazil (United States). We measure the forecast accuracy with the root mean squared forecast error (RMSFE). Also, the statistical significance of forecast accuracy improvements is assessed by implementing the Diebold and Mariano (1995) test where under the null hypothesis the two models have the same forecast accuracy. We conduct the DM test against a one–sided alternative hypothesis that factor models generate more accurate predictions than the benchmark. However, it is worth noting that these results should be treated with caution because the number of forecast errors for a given horizon is relatively small.

Figure 2.1 plots GDP growth against the first common factor based only on Google Trends categories for Brazil (left) and the United States (right) during the out–of–sample period. It is evident that for the United States the Google Trends factor tracks GDP growth rates quite well, while for Brazil it seems the factor is able to capture the main trends of GDP but with a slight delay. Therefore, it is reasonable to expect that Google Trends series are a more promising source of data for the United States rather than in Brazil.

Table 2.5 summarises the empirical results of the forecasting performance of dynamic factor models for Brazil and the United States. To quantify the forecast accuracy, we compare the RMSFE of factor models relative to a simple AR(1) benchmark. Subsequently, we will report comparisons between factor models with and without Google Trends categories. Results are reported for three monthly forecasts ($h = 1$), three monthly nowcasts ($h = 0$) and two (one) monthly backcasts ($h = -1$) for Brazil (United States). For each country, figures in bold specify the RMSFE–"best" models across all dimension–reduction models, for a given set of predictors and forecast horizon. Grey boxes indicate the RMSFE–"best" models across all groups of predictors and dimension–reduction methods, for a given forecast horizon and country.



(a) Brazil



(b) United States

Figure 2.1: Percent Change of GDP and Google Trends Common Factor

Notes: The left–hand (right–hand) plot shows the evolution of GDP growth rates (quarter–on–quarter) and the estimated first common factor based on Google Trends categories for Brazil (United States).

The results reveal several interesting insights. First, relative RMSFEs are consistently lower than unity for the United States, which implies that forecasts from DFMs are more accurate than those produced by the AR(1) benchmark during all prediction horizons; that is, they outperform, on average, by 12.8% in forecasting, by 18.8% in nowcasting, and by 21.8% in backcasting. For Brazil, factor models outperform the benchmark mostly in nowcasting and backcasting horizons, while for higher horizons, forecast gains tend to disappear. Specifically, factor models underperform the benchmark in forecasting on average by 10.7%, but they outperform in nowcasting and backcasting by 18.2% and 34.2%, respectively.

Second, the improvement in terms of forecast accuracy is noticeable both for Brazil and the United States, and the relative RMSFEs decrease as we move through the prediction period from the forecast to nowcast and backcast. This highlights the importance of updating the model with new information as we approach the date when GDP is published. It seems that in the case of Brazil, benefits from incorporating more information are more significant compared to the United States, since relative RMSFEs decrease at a higher rate, but this could be partially explained by the fact that for the United States factor models during forecasting horizons perform much better compared to Brazil. Also, the volatile nature of Brazilian GDP could weigh negatively on long–horizon forecasts.

Third, factor models that employ economic indicators and Google Trends categories tend to have the greatest performance against the benchmark for both countries, as shown by the grey boxes in Table 2.5. In particular, in the case of Brazil, in 5 out of 8 horizons the "globally–best" methods are those that employ both economic indicators and Google Trends categories. In the United States, this pattern is even more potent as in all horizons models that utilise both economic indicators and Google Trends categories generate the "globally–best" RMSFE results. In both cases, the LASSO approach produces the lowest relative RMSFE in the second and third round of forecasting and the first round of nowcasting, while the factor model without variable pre-selection reports the lowest relative forecast error in the final step of nowcasting and the first step of backcasting.

However, the variable selection method that produces the "locally–best" RMSFE results (see entries presented in bold in Table 2.5) is not consistent either across forecast horizons or across countries. Thus, it would be a tough assignment to select a particular dimension reduction approach *a priori*. Table 2.4 provides a summary regarding the RMSFE–best variable selection methods. The MRP approach fares quite well for Brazil by generating the lowest forecast error in 8 out of 24 cases (recall that we have eight horizons and three different initial sets of predictors), while in the United States the LASSO method wins in 9 out of 21 cases. In total, the LASSO attains the top rank in 16 cases out of 45.

Thus, our results are consistent with the majority of literature (Giannone et al., 2008; Bańbura et al., 2013) and highlight the strong performance of factor models relative to autoregressive models, especially in nowcasting and backcasting, as they can exploit more information each time new data are released. Also, findings imply that factor models based on both economic indicators and Google search categories generate significant forecast gains since, in most cases, the null hypothesis of equal forecast accuracy is being rejected (notice the plethora of numbers in Table 2.5 that are in grey boxes and signed with asterisks, meaning Diebold–Mariano test rejections) and produce the lowest forecast errors.

Finally, it is worth noting that the specification of the dynamic factor model and the decision on how to select the number of common factors might significantly affect the results. In the main set of findings, the IC_{p2} information criterion from Bai and Ng (2002) is employed. To enhance the generalizability of the results, we also utilise the IC_{p1} information criterion (see Table A.3). As can be observed from Table 2.5 and Table A.3, the results are relatively stable when using different information criteria.

2.4.3 Forecasting Performance of Factor Models with Targeted Predictors

To quantify the importance of including targeted predictors in a factor model, we compare the performance of factor models combined with variable selection methods to factor models without variable pre–selection. For example, when we use only economic indicators and Google Trends categories, each variable selection model (such as LASSO and Elastic Net) is compared to the factor model which uses all economic indicators and Google Trends categories.

Table 2.6 shows the corresponding relative RMSFEs and points to several interesting results. First, it is evident that forecast gains that arise from factor models with variable selection tend to decrease as we incorporate more information and the prediction horizon shortens; that is, relative RMSFEs tend to increase as we move from forecasting to nowcasting and backcasting. In particular, in Brazil's case, when only economic indicators are included, during the forecasting horizon, variable selection methods outperform the DFM without targeted predictors on average by 20.9%, while during the nowcasting and backcasting period underperform by 7.1% and 7.4%, respectively. Similarly, for the United States, during forecasting, variable selection methods outperform on average by 3.73%, while during nowcasting and backcasting, the simple factor model without variable pre–selection outperforms on average by 0.52% and 3.54%, respectively.

When Google Trends categories are included, variable selection methods outperform the factor models that include all predictors in forecasting horizons by 24.5% for Brazil and by 7.61% for the United States. However, in nowcasting and backcasting predictions, the factor model without pre–selection outperforms factor models with targeted predictors on average by 0.5% and 24.43%, respectively, for Brazil and by 2.13% and 9.15%, for the United States. For Brazil, only when Google Trends subcategories are incorporated, factor models with targeted predictors tend to perform better than a factor model without pre–selection during all horizons; that is, they outperform by 30.2% in forecasting, by 21.0% in nowcasting, and by 9.7% in backcasting. For the United States, when we add Google Trends subcategories, we have the same pattern as before, where variable selection methods outperform, on average, only in forecasting horizon, by 1.25%. In contrast, factor models without variable selection outperform factor models with pre-selection in nowcasting and backcasting horizons on average by 1.25% and 2.03%, respectively.

Interesting conclusions are derived if we analyse the individual performance of variable selection methods in factor models compared to those without variable pre–selection. In the case of Brazil, when only economic indicators are included in the model, the MRP approach outperforms in all horizons the factor model without variable pre–selection by 25.2%, 7.5%, and 15.7%, respectively, whereas, in the United States, the LASSO approach outperforms only in forecast and nowcast horizons by 6.5% and 2.5%, respectively.

When we add Google Trends categories, the LASSO works best for Brazil in forecasting and nowcasting and outperforms the factor model without variable selection by 34.8% and 9%, respectively, while it underperforms during backcasting by 5.1%. A similar pattern is observed in the United States, where the LASSO model outperforms the corresponding factor model without targeted predictors during forecasting and nowcasting, on average by 9.6% and 1.3%, whereas it underperforms in backcasting on average by 8.5%.

Finally, when Google Trends Subcategories are included, the MRP approach performs better for Brazil in forecasting and outperforms the factor model without targeted predictors on average by 34.3%, while the Elastic Net works best during nowcasting and backcasting and outperforms on average by 38.5% and 41.5%, respectively. For the

United States, the Elastic Net performs better in forecasting and nowcasting, where it outperforms on average by 2.4% and 0.7%, respectively, while in backcasting, the LASSO approach works best but produces almost the same performance compared to the factor model which includes all predictors.

Figure A.2 shows diagrammatically the forecast gains that arise when constructing factor models with targeted predictors. In particular, it shows the root mean squared forecast errors of the benchmark (AR(1) model) and the estimated factor models with and without variable pre–selection. As can be seen, the benchmark performs relatively well only in the case of Brazil at the $h = 1$ horizon and especially when economic indicators, Google Trends categories and subcategories are included in the model (Figure A.2e). The factor model estimated using all predictors outperforms the rest of the models mainly at backcast horizons (i.e., $h = -1$) for both countries. It is also evident that constructing targeted predictors before the computation of common factors generates some forecast gains at the forecast and early nowcast horizons in both cases. Moreover, the volatility of the forecast errors displays interesting insights as it seems that in the case of Brazil RMSFEs exhibit significantly higher volatility compared to the case of the United States in which forecast errors conducted from the models are much more stable.

Thus, benefits from pre–selecting predictors before constructing the factor model tend to arise mainly at one-quarter-ahead forecast horizons ($h = 1$), and the most significant gains from targeted predictors appear when we employ economic indicators, Google Trends categories, and subcategories in the case of Brazil, while the most significant gains for the United States arise when we utilise economic indicators and Google Trends categories. Also, when analysing the variable selection methods individually, the LASSO approach provides significant benefits when economic indicators and Google Trends categories are included, while the Elastic Net provides forecast improvements when economic indicators, Google Trends categories, and subcategories are included.

2.4.4 Forecasting Performance of Factor Models with Google Trends

In this section, we explicitly measure forecast gains that arise from the inclusion of Google Trends data by comparing the performance of each model specification that incorporates Google series (e.g., LASSO, Elastic Net) with the same method without Google series (e.g., LASSO, Elastic Net).

Table 2.7 directly quantifies the forecast benefits of including Google Trends data when the model already has access to information on economic conditions. For Brazil, when we include Google Trends categories, there are some gains at nowcasting and backcasting horizons. The LASSO approach also outperforms the corresponding LASSO approach without Google series during the second and the third step in forecasting. However, the value of Google data is mainly in backcasting, where it works well for all the methods other than ENET, although the improvements are not always statistically significant. Interestingly, when we add Google Trends subcategories, forecast gains tend to disappear during all prediction horizons, except in the elastic net method (ENET), which improves the model in nowcasting and backcasting horizons. On average, factor models based on Google Trends subcategories underperform in all horizons by 17.5%, 21.3%, and 15.8%, respectively.

Moving to the United States, a similar pattern is observed. Forecast gains appear only when Google Trends categories are included, while when we add the subcategories, forecast benefits vanish. Specifically, models that incorporate Google Trends categories outperform the corresponding models without Google series during forecasting, nowcasting, and backcasting on average by 3.1%, 5.1%, and 5.4%, respectively. On the other hand, models that employ both Google categories and subcategories underperform their benchmarks by 1.9%, 4.6%, and 5.5%, respectively. For the U.S., the relative gains are the greatest for MRP across all horizons from including Google Trends, although MRP is only statistically significantly better at the longest forecast horizon.

Figure A.3 shows graphically the RMSFEs of the three factor models that employ economic indicators, economic indicators and Google search categories, and economic indicators, Google search categories and subcategories. In almost all cases the Google-based factor models produce a lower forecast error compared to the factor models that exclude Google search data as shown by the dashed lines. The only exception is in the case of Brazil in one-quarter-ahead forecasts and early nowcasts.

Figure A.4 attempts to characterise the common factors by showing the coefficient of determination (R^2) of the regressions of individual predictors against each of the three common factors over the full sample period. The individual predictors are grouped by category. Generally speaking, the first factor for Brazil is related with *Prices* while for the U.S. with *Leading* and *Survey* indicators. The second factor loads on *Economic* and *Retail* indicators in both countries, but factor loadings spread out also to other categories. Finally, the third factor reflects primarily *Google Trends Categories*, meaning that Google search series have indeed a significant presence in the factor model.

Overall, forecast gains when including Google data appear when we incorporate only broad Google Trends categories, while when we include the subcategories, forecast gains vanish. This suggests that subcategories might repackage information already captured in the main Google Trends categories, and any further disaggregation does not appear to improve forecast performance. Also, on average, Google data seems to improve the performance of factor models to a similar extent in the United States and in Brazil. For both countries, the improvements depend on the variable–selection strategy underpinning the factor model, and are not always statistically significant.

A priori, one might suppose Google data would be more informative about consumer behaviour when the level of *discretionary* consumption is higher, and more consumers use Google, which would favour the United States. Discretionary consumption and internet penetration rates are both higher in the United States: U.S. consumers are more likely to use Google searches to inform their decisions. However, the U.S. also has high–quality alternative sources of information – the economic indicators, which are valuable in predicting the course of the economy. We hypothesise that Google data might have an edge for predicting specific categories of expenditure, such as consumer durable expenditure,

Table 2.4: Summary of the RMSFE–best Variable Selection Methods

	All	LASSO	AdaLASSO	ENET	MRP
		Brazil			
Model 1	2	2	0	0	4
Model 2	2	5	0	0	1
Model 3	0	0	0	5	3
Total-BR	4	7	0	5	8
		United States			
Model 1	1	4	0	1	1
Model 2	2	3	0	1	1
Model 3	1	2	1	3	0
Total-US	4	9	1	5	2
Total	**8**	**16**	**1**	**10**	**10**

Notes: *Model 1* refers to factor models based on Economic Indicators, *Model 2* denotes models based on Economic Indicators and Google Trends Categories, and *Model 3* represents specifications based on Economic Indicators, Google Trends Categories and Subcategories.

although we do not consider this here.

Finally, although our sample of forecast errors is necessarily short (because of the availability of Google Trends data), we provide some rudimentary analysis of whether forecast performance changes over time. Table Table A.1 and Table A.2 in Appendix A split the forecast evaluation period into two: 2014–2016 and 2017–2019, and show broadly similar patterns of results to Table 2.7. For example, Google Trends generate relative improvements on both sub–periods using LASSO for Brazil.

2.4.5 The Selected Predictors and Google Trends Categories

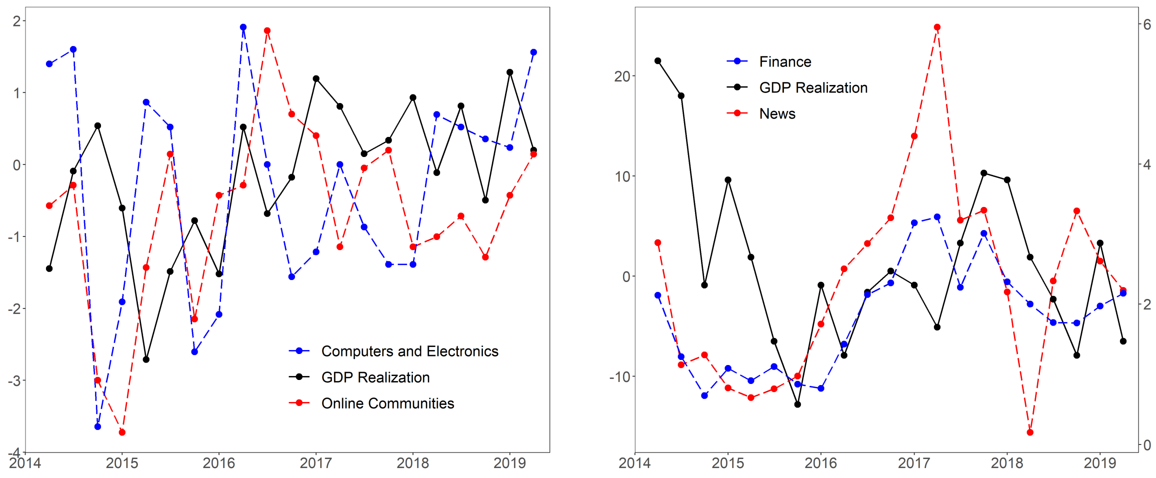
In the final part of our analysis, we focus on the predictors selected by the "most recent performance" method. Figure A.5 shows the importance of each variable group for a given model for each country. The values in the graphs are scaled such that they sum to one. A first glimpse at the graphs reveals that when only economic indicators are included, the composition of predictors remains relatively stable during the out–of–sample period. However, it is clear that when Google search series are incorporated, the share of each group of predictors becomes quite volatile, especially when all Google search series are included, i.e., Google Trends categories and subcategories.

It is worth mentioning, that in the case of Brazil, the dataset with targeted predictors is being updated 23 times through the out–of–sample period, while in the case of the United States the dataset is being updated 45 times. The difference in the number of times the datasets are being updated is explained by the fact that the publication delay for Brazilian GDP (around eight weeks) is two times larger than the publication delay of the United States GDP (around four weeks). Thus, in the United States, we have more frequently an updated balanced dataset and can re–estimate the variable selection methods more often.

Moreover, interesting insights are gained when analysing the individual Google Search categories that are being selected by the MRP method. Table 2.8 exhibits the 10 most frequent Google series for each country. Somewhat surprisingly, the most frequently selected category for Brazil is Online Communities, selected in 96% of cases and includes queries related to community websites and Social Networks. The next most frequently chosen categories are Computers & Electronics and Business & Industrial which seem logical to be connected with the overall economic activity, as they incorporate queries such as computer hardware, consumer electronics, agriculture, and construction.

Turning to the United States, the most frequently selected category is News which includes queries such as business news, local and world news, and politics, and has been selected in all cases. Next, Finance is selected in 96% of cases and incorporates search terms that are directly associated with the state of the economy such as investing, lending, and insurance. Finally, the third most frequently selected category is Home & Garden which contains searches related to home improvement, appliances, and furnishings. Figure 2.2 shows the evolution of GDP growth rates and the two most frequently selected Google search categories for each country.

It is noteworthy that only three categories are common in both lists with categories that are most frequently selected by the MRP method: Sports, Finance, and Home & Garden. Clearly, these categories are related to the services sector. It is reasonable to expect that the majority of the Google Trends categories as well as the common categories to be connected with services since the tertiary sector contributes the most in



(a) Brazil (b) United States

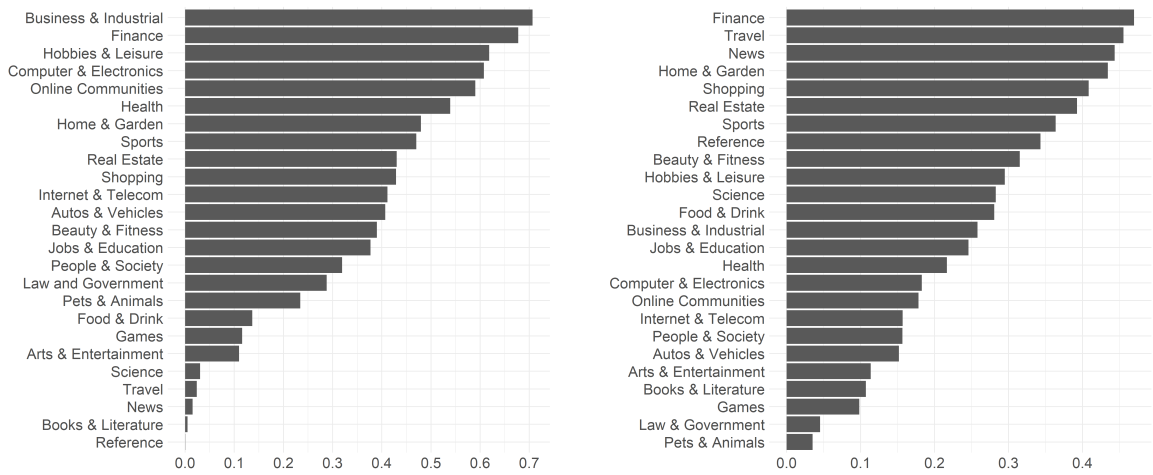
Figure 2.2: Percent Change of GDP and Google Trends Categories

Notes: The left–hand (right–hand) plot shows the evolution of GDP growth rates (quarter–on–quarter) and the two most frequent Google Trends categories selected by the MRP method for Brazil (United States).

both countries[9]. The fact that in the case of Brazil there are a few categories related to the manufacturing sector highlights the structural differences between the economies. Brazil is heavily dependent on the primary and secondary sectors and this can be reflected from the frequent usage of categories like Computers & Electronics and Business & Industrial. On the other hand, the U.S. economy has a relatively larger services sector, and intuitively search terms related to News and Finance are the most frequently used.

Finally, Figure 2.3 shows the contribution of each Google Trends category to the first Google Trends common factor by regressing the common factor against each Google category and then extracting the R^2. Categories like Business & Industrial, Finance and Hobbies & Leisure seem to have a significant contribution to the Brazilian Google Trends factor, while categories like Finance, Travel, and News appear to have a substantial contribution to the U.S. common factor.

[9]The services sector consists of around 77% of GDP in the United States and 63% in Brazil (Statista, 2018).



(a) Brazil **(b)** United States

Figure 2.3: The Contribution of Google Trends Categories to the Google Factor

Notes: The left–hand (right–hand) plot shows the contribution of Google Trends categories to the first Google Trends common factor computed for Brazil (United States). The contribution of each variable is measured by the coefficient of determination (R^2) extracted from regressions of the first Google Trends common factor against the individual Google Trends categories over the entire sample.

2.5 Concluding Remarks

Although many studies investigate the potential usefulness of "Big Data" for forecasting specific macroeconomic variables such as unemployment and inflation, only a few focus on overall economic activity – GDP. This chapter contributes to the literature by exploring whether Google search data complements more traditional economic indicators to provide forecast gains in a nowcasting exercise.

In a pseudo–real–time framework we estimate dynamic factor models to nowcast GDP growth rates for Brazil and the United States, from 2014 to 2019. We consider the efficacy of several variable selection methods, including the elastic net, least absolute shrinkage and selection operator (LASSO), and an adaptive version of LASSO. This provides evidence on whether forecast gains arise from estimating factor models on targeted predictors. Additionally, we construct sets of targeted predictors from what we call the "most recent performance" approach. Every time the balanced dataset of predictors is updated with new information we choose the variable selection method that produces the "best" local out–of–sample performance.

There are a number of findings. Firstly, factor models effectively incorporate the new information which is published within the reference quarter. Factor models outperform the AR(1) benchmark at all horizons for the United States, while for Brazil, they outperform the benchmark mostly at nowcasting and backcasting horizons.

Secondly, we find factor models that utilise both economic indicators and Google Trends categories have the best performance against the autoregressive model. Variable selection methods work best at forecast horizons, with diminishing performance as additional information is included (now– and back–casting). Perhaps unsurprisingly, the largest gains from variable selection occur when both Google data and economic indicators are allowed, rather than when the information set is restricted to economic indicators.

Thirdly, only the main Google Trends categories are found to have the potential to predict GDP growth rates, since when we add to the model the subcategories forecast gains vanish. This suggests that the information contained in subcategories is already captured by the main categories.

Finally, our results do not clearly suggest Google data are more valuable for one country than the other. For both countries, their value depends on the chosen variable–selection method.

Although our main conclusions regarding the overall usefulness of Google Trends data in forecasting economic activity are in line with Ferrara and Simoni (2019) and Götz and Knetsch (2019), they contrast with the former study with respect to the horizons at which Google search series are useful. However, our results corroborate the findings of both papers concerning the benefits of a variable–selection step for Google Trends data.

Finally, there are potential limitations and extensions. First, our pseudo–real–time framework. Although data revisions might only have a small effect on the computation of common factors, the use of real–time vintages of data for GDP growth may affect the findings. However, Bernanke and Boivin (2003) and Clements (2016) find that the relative forecasting performance of factor models and AR models in real–time and pseudo out–of–sample exercises is similar. Nevertheless, this may depend on the nature of the revisions: see Clements and Galvão (2019) for a recent review of data revisions and forecasting. A possible extension would be to consider the conceptually distinct effects of model re–estimation and the newly–released data points when a nowcast is revised as more data becomes available. This exercise might usefully be disaggregated across different blocks of variables.

Table 2.5: Forecast Evaluation of Factor Models (IC_{p2})

	Forecast ($h = 1$)			Nowcast ($h = 0$)			Backcast ($h = -1$)	
	1	2	3	1	2	3	1	2
Brazil								
Economic Indicators								
All	1.44	1.31	1.24	1.06	**0.72****	**0.55****	0.65*	0.59*
LASSO	**1.06**	1.04	**0.90***	0.94	0.91	0.68**	0.71*	0.72*
AdaLASSO	1.19	1.13	0.99	1.02	1.00	0.77**	0.78*	0.81*
ENET	1.22	1.11	0.98	0.98	0.85*	0.65**	0.64*	0.62*
MRP	1.07	**0.95***	0.97	**0.87***	0.73**	0.55**	**0.53***	**0.52***
Economic Indicators and Google Trends categories								
All	1.52	1.40	1.37	1.23	0.80**	**0.46****	**0.49***	0.52*
LASSO	**1.07**	**0.88**	**0.86***	**0.90***	**0.74***	0.63**	0.65*	0.61*
AdaLASSO	1.29	1.28	1.03	1.11	1.04	0.68**	0.72*	0.72*
ENET	1.22	1.09	1.01	1.02	0.89*	0.70**	0.68*	0.65*
MRP	1.21	1.02	1.00	0.99	0.79**	0.54**	0.49*	**0.50***
Economic Indicators, Google Trends categories and subcategories								
All	1.82	1.85	1.63	1.53	1.33	0.89	0.88	0.81
LASSO	1.26	1.22	1.14	1.17	1.09	0.83	0.88	0.87
AdaLASSO	1.45	1.45	1.28	1.47	1.37	1.04	1.17	1.09
ENET	1.28	1.17	**1.09**	0.99	**0.76****	**0.56****	**0.52***	**0.48***
MRP	**1.18**	**1.15**	1.15	**0.98**	1.03	0.56**	0.57*	0.55*
United States								
Economic Indicators								
All	0.86	0.93	0.91*	0.83	0.84	0.76	**0.76**	
LASSO	0.80	**0.88**	**0.84****	**0.81**	0.80*	**0.75***	0.77	
AdaLASSO	0.84	0.90	0.94	0.91	0.86	0.83	0.81	
ENET	**0.79**	0.91	0.86*	0.82	0.81*	0.76*	0.76	
MRP	0.84	0.90*	0.89*	0.83	**0.78***	0.79*	0.79	
Economic Indicators and Google Trends categories								
All	0.86	0.95	0.90*	0.83	0.77*	**0.68***	**0.68***	
LASSO	0.80	**0.84***	**0.81***	**0.76***	0.76*	0.73*	0.74	
AdaLASSO	0.81	0.87	0.89*	0.87	0.81*	0.77*	0.77	
ENET	**0.78**	0.85*	0.83*	0.79	0.80*	0.76*	0.74	
MRP	0.79	0.91	0.82*	0.77*	**0.75***	0.75*	0.72	
Economic Indicators, Google Trends categories and subcategories								
All	0.84	0.92	0.85*	0.81	0.81	0.79*	**0.77**	
LASSO	**0.83**	0.89	0.86*	0.82	0.81	**0.78***	0.77	
AdaLASSO	0.85	0.89	**0.83****	0.85	0.82*	0.82*	0.80	
ENET	0.83	**0.89**	0.83**	**0.80***	**0.80***	0.79*	0.79	
MRP	0.83	0.91	0.87*	0.81	0.83	0.81	0.78	

Notes: Entries in this table show Root Mean Squared Forecast Errors (RMSFEs) of Dynamic Factor Models (DFMs) relative to the AR(1) benchmark. Therefore, entries lower than one suggest that a particular prediction based on a DFM is more accurate compared to the AR(1) benchmark. Numbers in bold indicate predictions with the lowest relative RMSFE for each forecast horizon, within a given DFM specification. Grey boxes denote the lowest error measure achieved by the competing models for a given country and forecast horizon. Numbers followed by asterisks (**5% level, *10% level) are significantly superior to the AR(1) benchmark based on the Diebold–Mariano test.

Table 2.6: Forecast Evaluation of Factor Models with Targeted Predictors

	Forecast ($h = 1$)			Nowcast ($h = 0$)			Backcast ($h = -1$)	
	1	2	3	1	2	3	1	2
				Brazil				
Economic Indicators								
LASSO	**0.74**	0.79	**0.73**	0.89	1.27	1.23	1.10	1.20
AdaLASSO	0.82	0.87	0.80	0.96	1.40	1.40	1.21	1.37
ENET	0.85	0.85	0.79	0.92	1.19	1.19	0.99	1.04
MRP	0.74	**0.72**	0.78	**0.82**	**1.02**	**1.00**	**0.82**	**0.87**
Economic Indicators and Google Trends categories								
LASSO	**0.70**	**0.63**	**0.63**	**0.73**	**0.92**	1.36	1.33	1.18
AdaLASSO	0.85	0.92	0.75	0.90	1.30	1.48	1.48	1.37
ENET	0.80	0.78	0.74	0.83	1.10	1.52	1.40	1.24
MRP	0.79	0.73	0.73	0.80	0.98	**1.17**	**1.01**	**0.97**
Economic Indicators, Google Trends categories and subcategories								
LASSO	0.69	0.66	0.70	0.77	0.82	0.94	1.00	1.07
AdaLASSO	0.79	0.78	0.79	0.96	1.03	1.17	1.33	1.34
ENET	0.70	0.63	**0.67**	0.64	**0.58**	**0.63**	**0.58**	**0.59**
MRP	**0.65**	**0.62**	0.71	**0.64**	0.77	0.63	0.65	0.67
				United States				
Economic Indicators								
LASSO	0.94	**0.94**	**0.93**	**0.98**	0.96	**0.99**	1.02	
AdaLASSO	0.98	0.96	1.03	1.10	1.03	1.08	1.08	
ENET	**0.92**	0.97	0.95	0.99	0.97	1.00	**1.01**	
MRP	0.98	0.96	0.98	1.00	**0.94**	1.04	1.04	
Economic Indicators and Google Trends categories								
LASSO	0.93	**0.88**	**0.90**	**0.92**	0.99	**1.07**	1.09	
AdaLASSO	0.94	0.91	1.00	1.05	1.05	1.13	1.13	
ENET	**0.91**	0.89	0.93	0.95	1.04	1.10	1.09	
MRP	0.92	0.96	0.92	0.93	**0.98**	1.09	**1.06**	
Economic Indicators, Google Trends categories and subcategories								
LASSO	**0.98**	0.97	1.01	1.01	1.00	**0.99**	**1.00**	
AdaLASSO	1.01	0.97	**0.98**	1.05	1.01	1.04	1.03	
ENET	0.99	**0.97**	0.98	**1.00**	**0.99**	1.00	1.03	
MRP	0.99	0.99	1.02	1.00	1.03	1.03	1.02	

Notes: Entries in this table denote relative Root Mean Squared Forecast Errors (RMSFEs) of Dynamic Factor Models (DFMs) based on different variable selection methods to RMSFEs of a factor model without a variable pre-selection. Therefore, entries lower than one suggest that a particular factor model with targeted predictors is more accurate compared to a factor model without variable pre-selection, for a given forecast horizon. Numbers in bold indicate model specifications with the “locally–best” RMSFEs, for a given forecast horizon.

Table 2.7: Forecast Evaluation of Factor Models with Google Trends

	Forecast ($h = 1$)			Nowcast ($h = 0$)			Backcast ($h = -1$)	
	1	2	3	1	2	3	1	2
				Brazil				
Economic Indicators and Google Trends categories								
All	1.06	1.07	1.10	1.16	1.12	**0.84***	**0.75***	0.88
LASSO	1.00	**0.84***	**0.95**	**0.96**	**0.82***	0.93	0.90	**0.86***
AdaLASSO	1.09	1.13	1.04	1.09	1.04	0.89	0.92	0.88
ENET	**1.00**	0.98	1.03	1.04	1.04	1.08	1.05	1.05
MRP	1.13	1.08	1.03	1.14	1.08	0.98	0.92	0.98
Economic Indicators, Google Trends categories and subcategories								
All	1.27	1.42	1.32	1.45	1.85	1.62	1.36	1.37
LASSO	1.19	1.17	1.26	1.24	1.20	1.23	1.23	1.21
AdaLASSO	1.22	1.28	1.29	1.44	1.36	1.36	1.49	1.34
ENET	**1.04**	**1.05**	**1.11**	**1.01**	**0.90**	**0.85***	**0.80***	**0.77***
MRP	1.10	1.22	1.19	1.12	1.40	1.02	1.07	1.06
				United States				
Economic Indicators and Google Trends categories								
All	1.04	1.02	0.99	1.00	**0.92**	**0.90**	**0.90**	
LASSO	0.96	0.94	0.97	**0.94**	0.95	1.00	0.98	
AdaLASSO	0.94	0.97	0.98	0.95	0.94	0.93	0.94	
ENET	1.00	**0.93**	0.96	1.04	1.01	1.02	1.01	
MRP	**0.91**	0.94	**0.89***	0.96	0.96	0.94	0.92	
Economic Indicators, Google Trends categories and subcategories								
All	1.04	0.98	**0.93***	**0.97**	**0.97**	1.03	1.02	
LASSO	1.02	1.00	1.01	1.11	1.06	1.07	1.04	
AdaLASSO	**0.98**	1.00	1.04	1.05	1.02	**0.99**	**1.01**	
ENET	1.10	**0.98**	1.00	1.11	1.06	1.05	1.04	
MRP	1.04	1.00	1.05	1.23	1.23	1.03	1.06	

Notes: Entries in this table denote relative Root Mean Squared Forecast Errors (RMSFEs) of factor models constructed with Economic Indicators and Google search series compared to factor models based only on Economic Indicators. Therefore, entries lower than one suggest that a particular DFM specification that utilises Google data next to economic indicators is more accurate compared to factor models based on economic indicators of the same variable selection method, for a given forecast horizon. Numbers in bold indicate model specifications with the “locally–best” RMSFEs, for a given forecast horizon. Grey boxes denote the lowest error measure achieved by the competing models for a given country and forecast horizon. Numbers followed by asterisks (**5% level, *10% level) are significantly superior to the DFM based on only economic indicators according to the Diebold–Mariano test.

Table 2.8: Google Trends Categories Selected

Brazil		United States	
Category	Frequency	Category	Frequency
Online Communities	96%	News	100%
Computers & Electronics	91%	**Finance**	**96%**
Business & Industrial	91%	**Home & Garden**	**96%**
Sports	**87%**	Reference	93%
Health	87%	Jobs & Education	91%
Finance	**83%**	**Sports**	**89%**
Internet & Telecom	83%	Science	89%
Games	74%	Hobbies & Leisure	87%
Home & Garden	**74%**	Food & Drink	87%
Shopping	74%	Beauty & Fitness	82%

Notes: Bold entries denote common Google Trends categories for both countries.

A Appendix

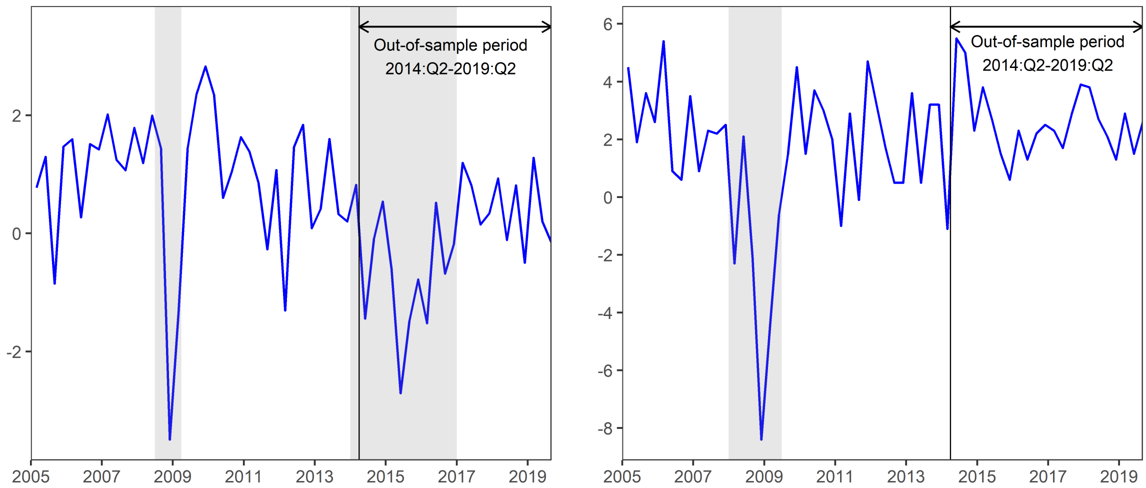
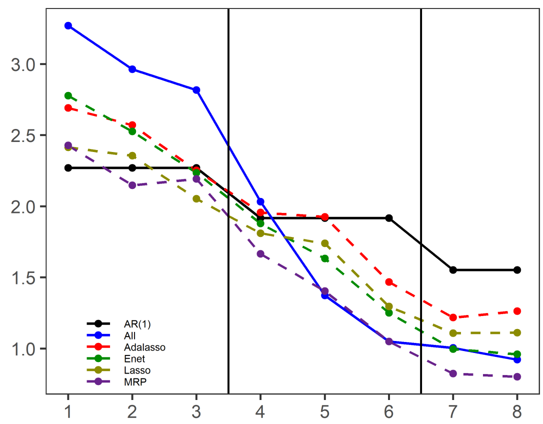
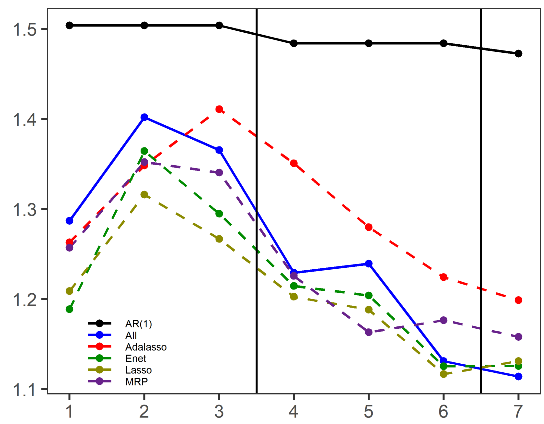


Figure A.1: Percent Change of GDP of Brazil (Left) and the US (Right)

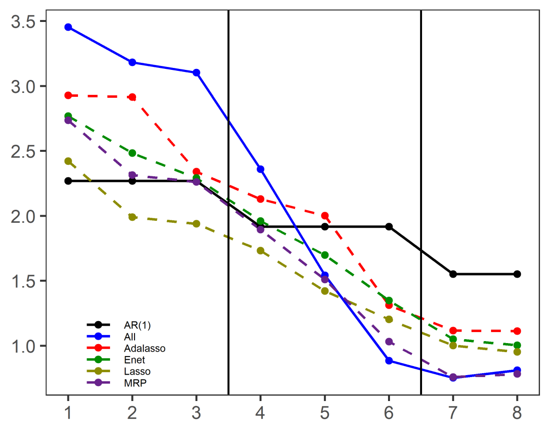
Notes: The left–hand plot shows the evolution of GDP growth rates (year–over–year%) for Brazil, from January 2005 to September 2019. Shaded areas indicate periods of recessions. The Brazilian economy suffered two recessionary periods during the sample period; from the third quarter of 2008 until the first quarter of the next year, and from the first quarter of 2014 until the first quarter of 2017, respectively. The right–hand plot presents GDP growth rates for the United States. The US economy fell into a recessionary territory from the first quarter of 2008 until the third quarter of 2009.



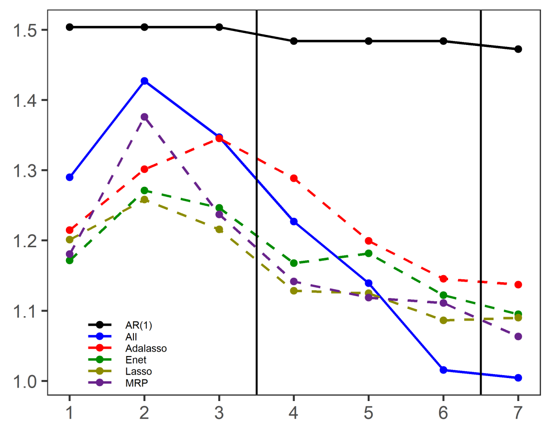
(a) Economic Indicators – BR



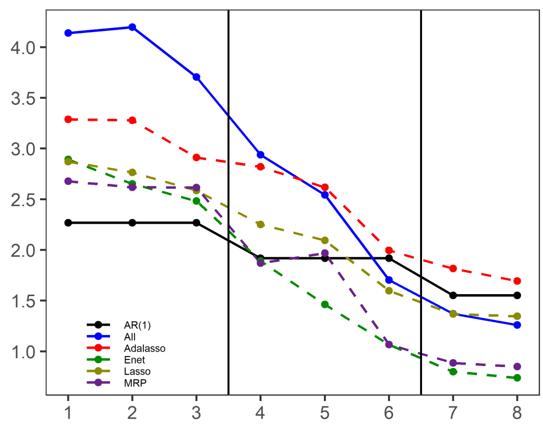
(b) Economic Indicators – US



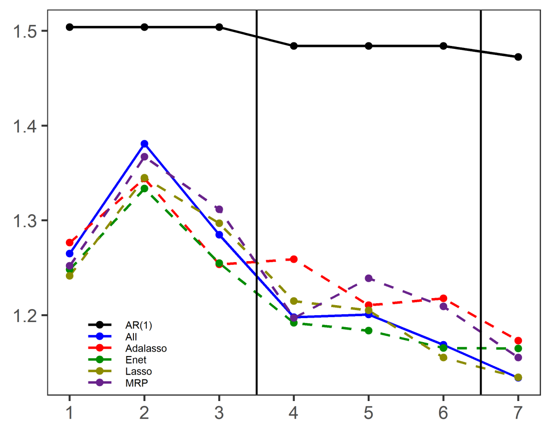
(c) Economic Indicators, GTs Categories – BR



(d) Economic Indicators, GTs Categories – US

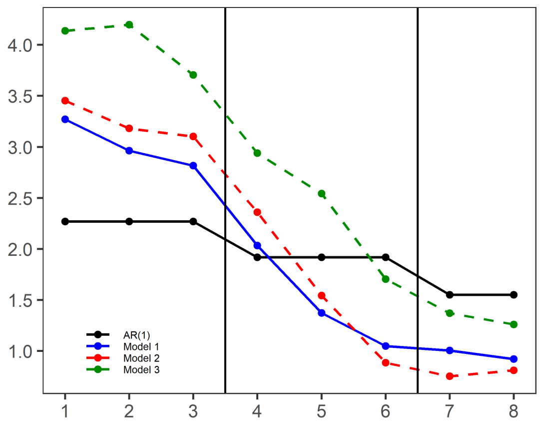


(e) Economic Indicators, GTs Categories & Subcategories – BR

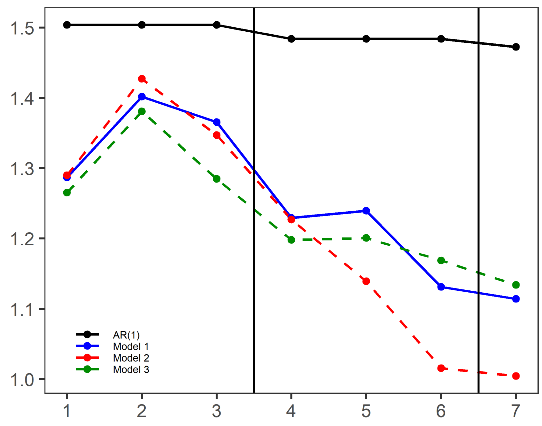


(f) Economic Indicators, GTs Categories & Subcategories – US

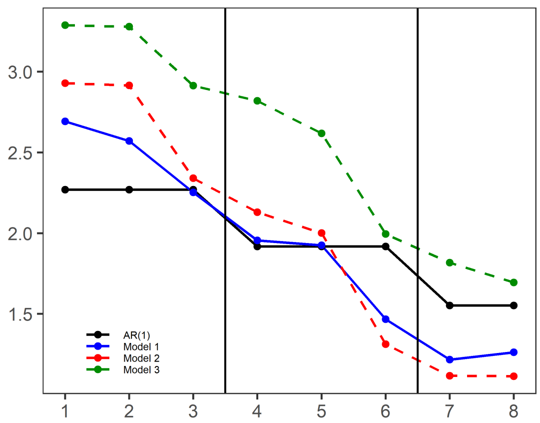
Figure A.2: Does Variable Selection Matter? (RMSFE)



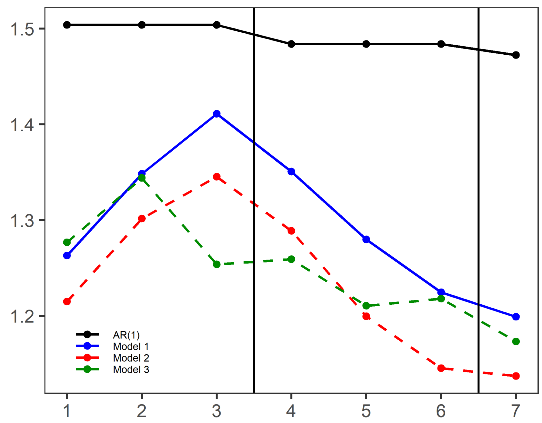
(a) All Predictors – BR



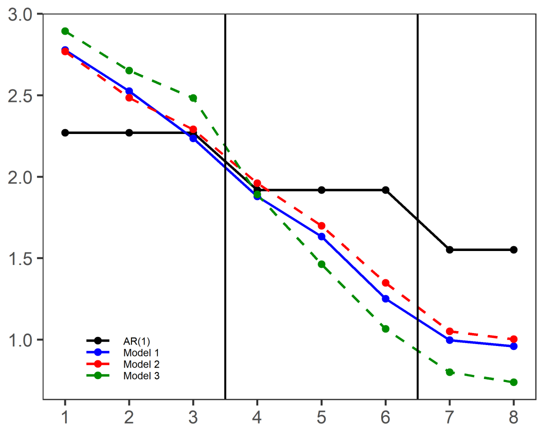
(b) All Predictors – US



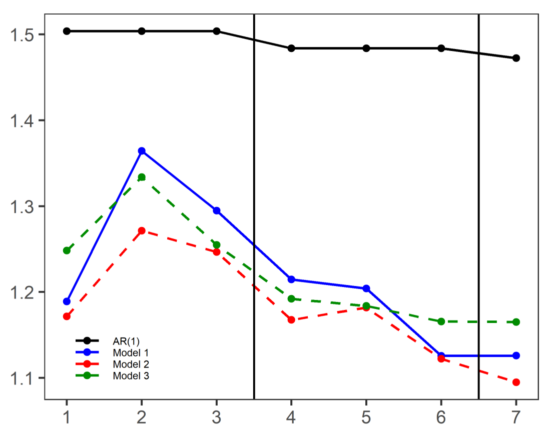
(c) Targeted Predictors, AdaLASSO – BR



(d) Targeted Predictors, AdaLASSO – US

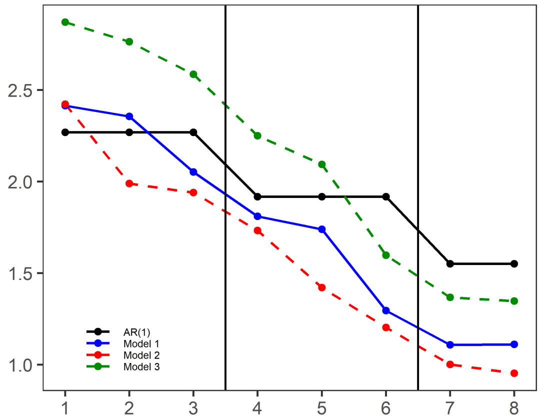


(e) Targeted Predictors, ENET – BR

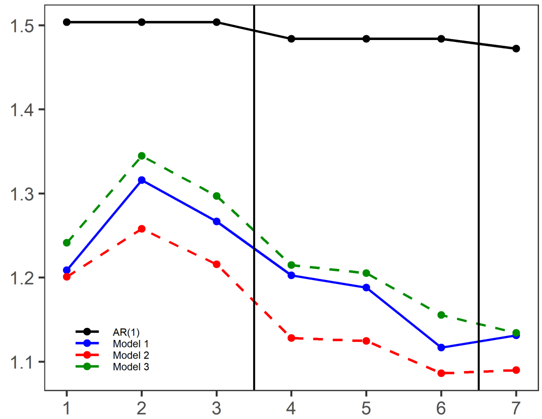


(f) Targeted Predictors, ENET – US

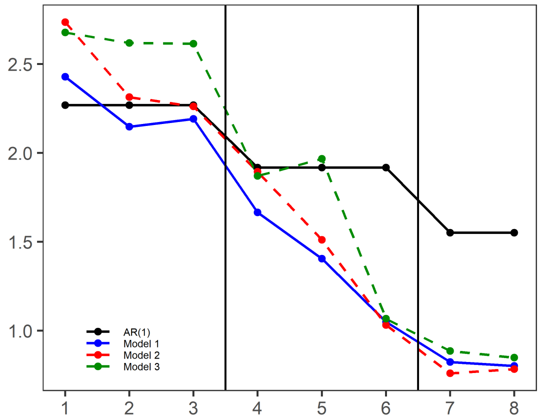
Figure A.3: Do Google Trends Data Provide Forecast Gains? (RMSFE)



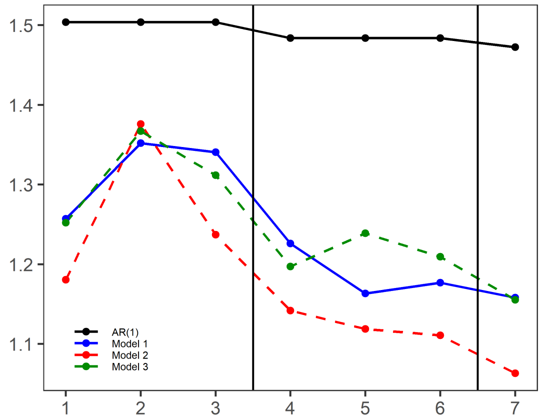
(g) Targeted Predictors, LASSO – BR



(h) Targeted Predictors, LASSO – US



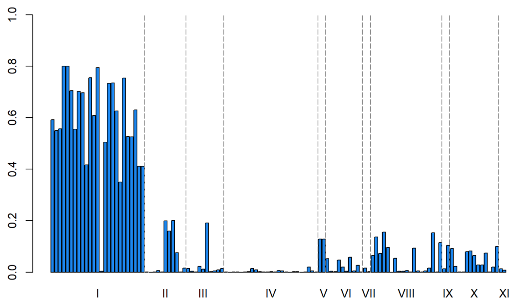
(i) Targeted Predictors, MRP – BR



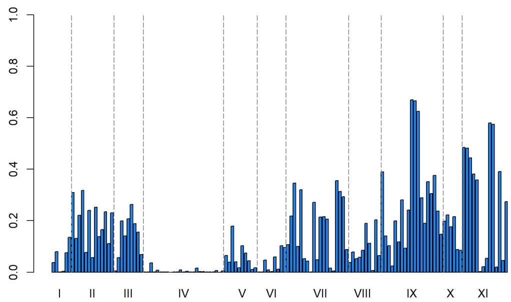
(j) Targeted Predictors, MRP – US

Figure A.3: Do Google Trends Data Provide Forecast Gains? (RMSFE) (cont.)

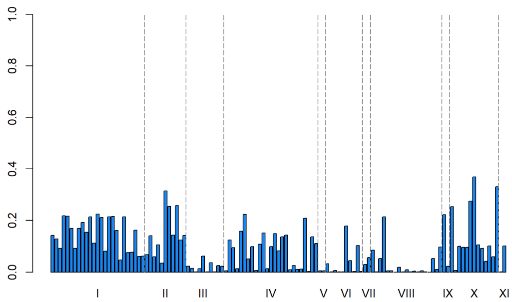
Notes: *Model 1* refers to dynamic factor models based on Economic Indicators, *Model 2* refers to factor models based on Economic Indicators and Google Trends Categories, and *Model 3* represents factor models based on Economic Indicators, Google Trends Categories and Subcategories. Left–hand plots represent factor model RMSFE for Brazil, while right–hand plots show RMSFE for the United States. Horizontal axes show the timing of monthly predictions produced for each reference quarter: steps 1 to 3 represent one–quarter–ahead *forecasts*, steps 4 to 6 denote current quarter *nowcasts*, and the rest steps represent forecasts for the preceding quarter, i.e., *backcasts*.



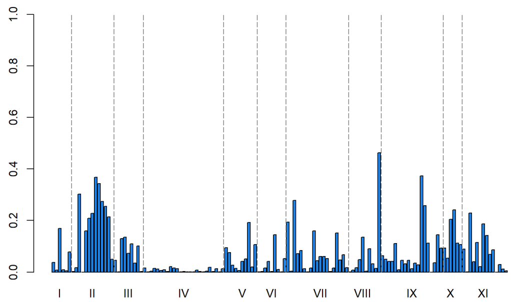
(a) First Common Factor – BR



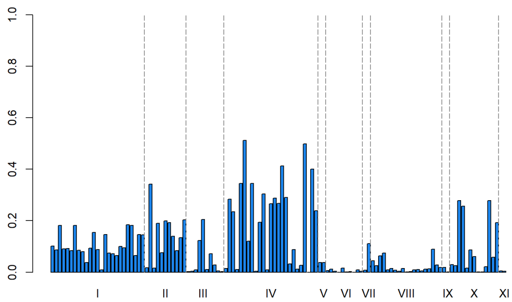
(b) First Common Factor – US



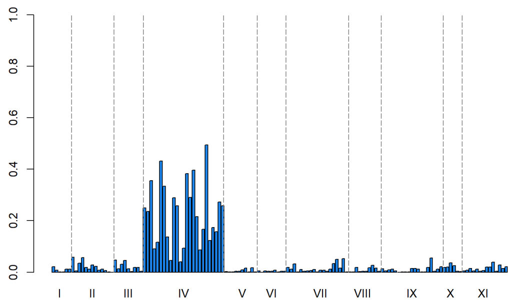
(c) Second Common Factor – BR



(d) Second Common Factor – US



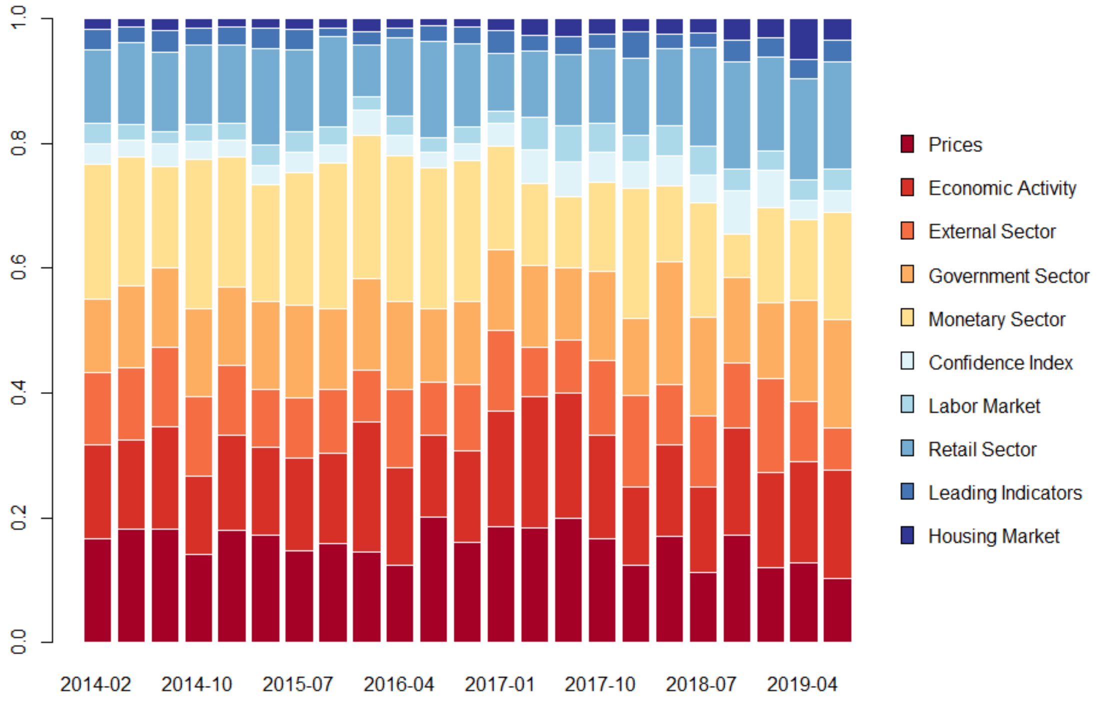
(e) Third Common Factor – BR



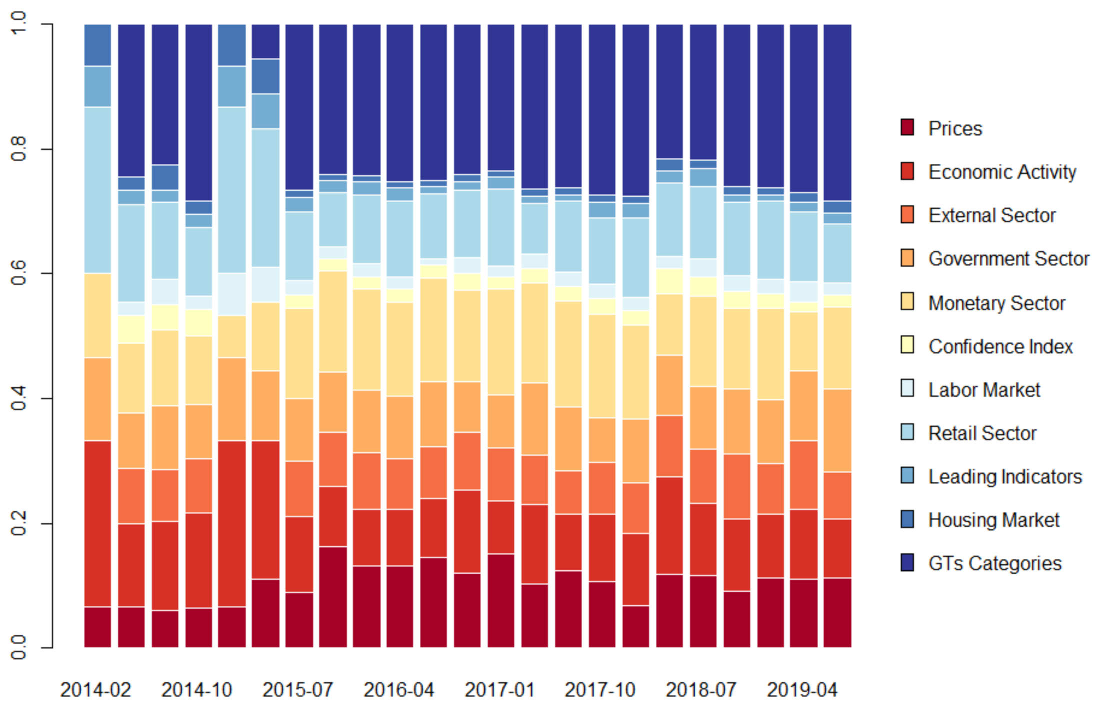
(f) Third Common Factor – US

Figure A.4: R–squared Between the Individual Predictors and the Common Factors

Notes: These figures show the coefficient of determination (R^2) obtained from simple regressions of the individual predictors (one at a time) on each of the three common factors over the entire sample. Left–hand plots present the R^2 for Brazil, while right–hand plots show the R^2 for the United States. *Prices* are denoted by group I, *Economic Activity* indicators by II, *External Sector* by III, *Google Trends Categories* by IV, *Housing Market* by V, *Government Sector* by VI, *Labor Market* by VII, *Monetary Sector* by VIII, *Leading Indicators* by IX, *Retail Sector* by X, *Survey Indicators* by XI.

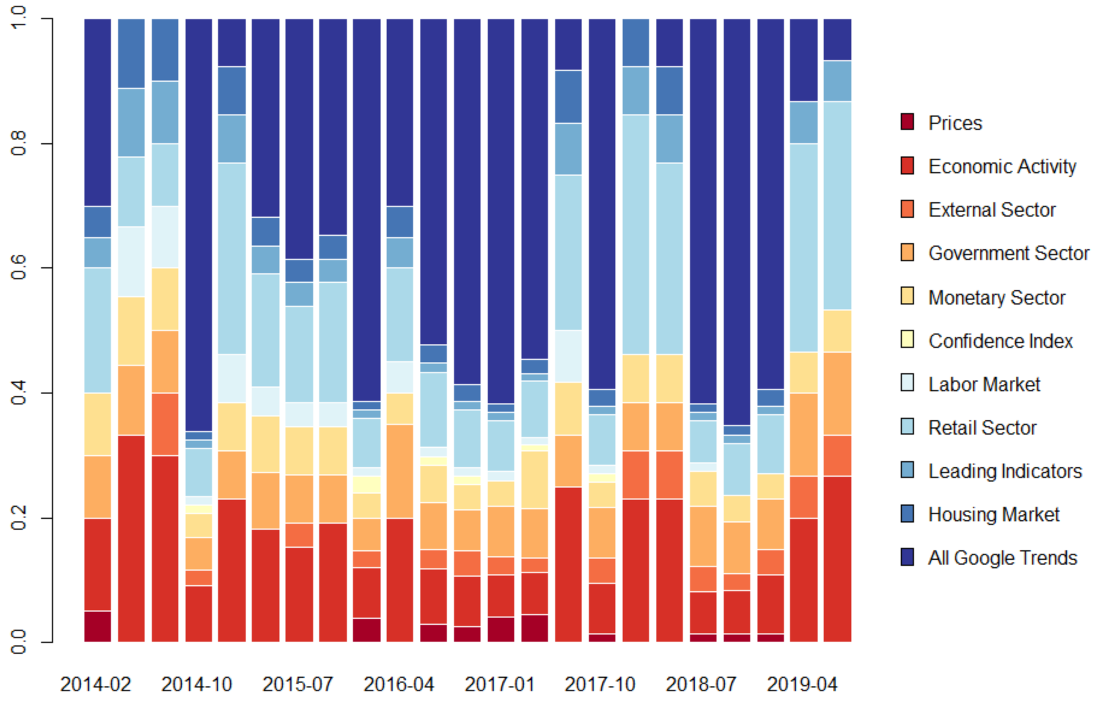


(a) Economic Indicators, Brazil

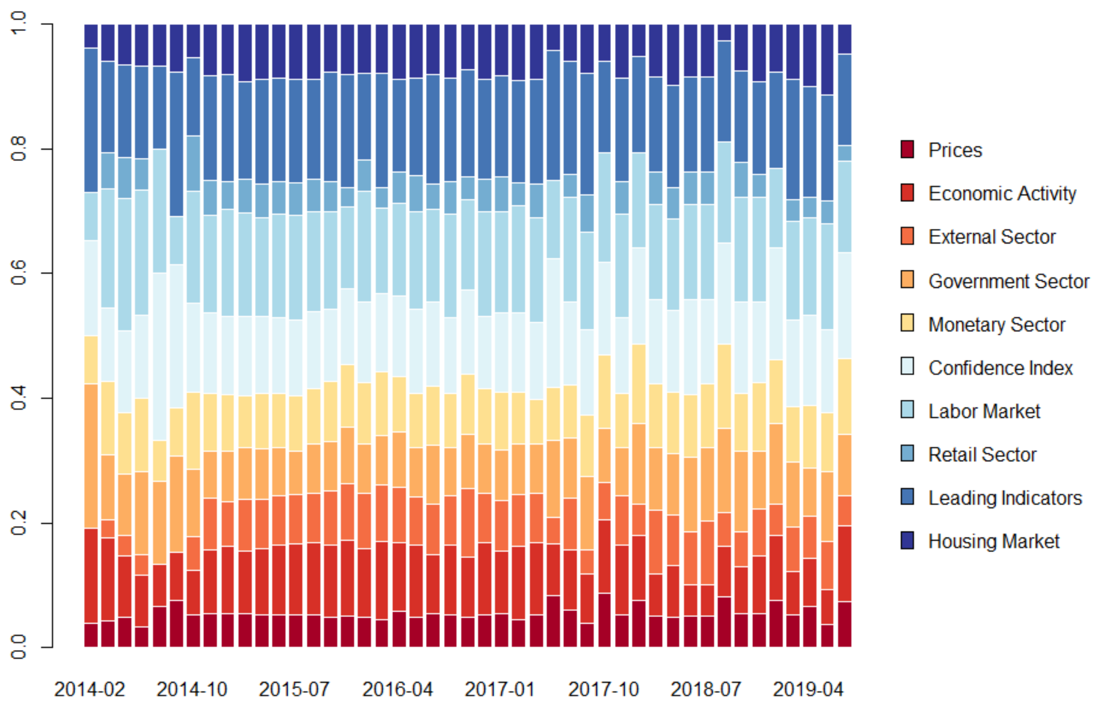


(b) Economic Indicators and Google Trends categories, Brazil

Figure A.5: Variable Selection in MRP Approach

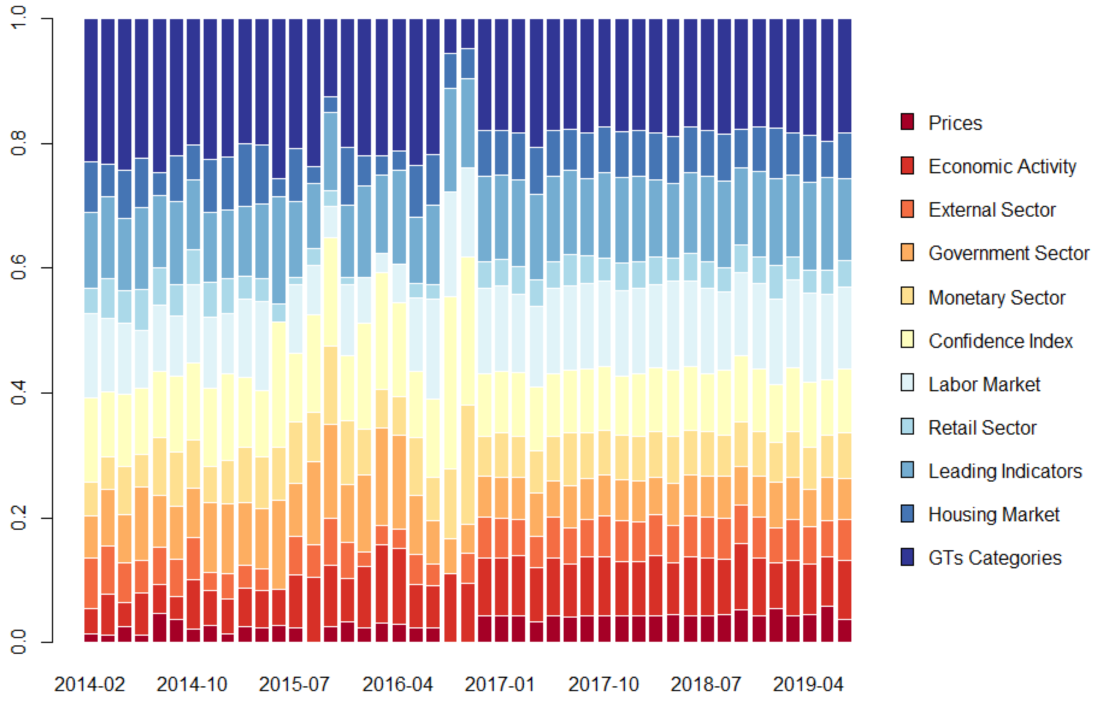


(c) Economic Indicators, Google Trends categories and subcategories, Brazil

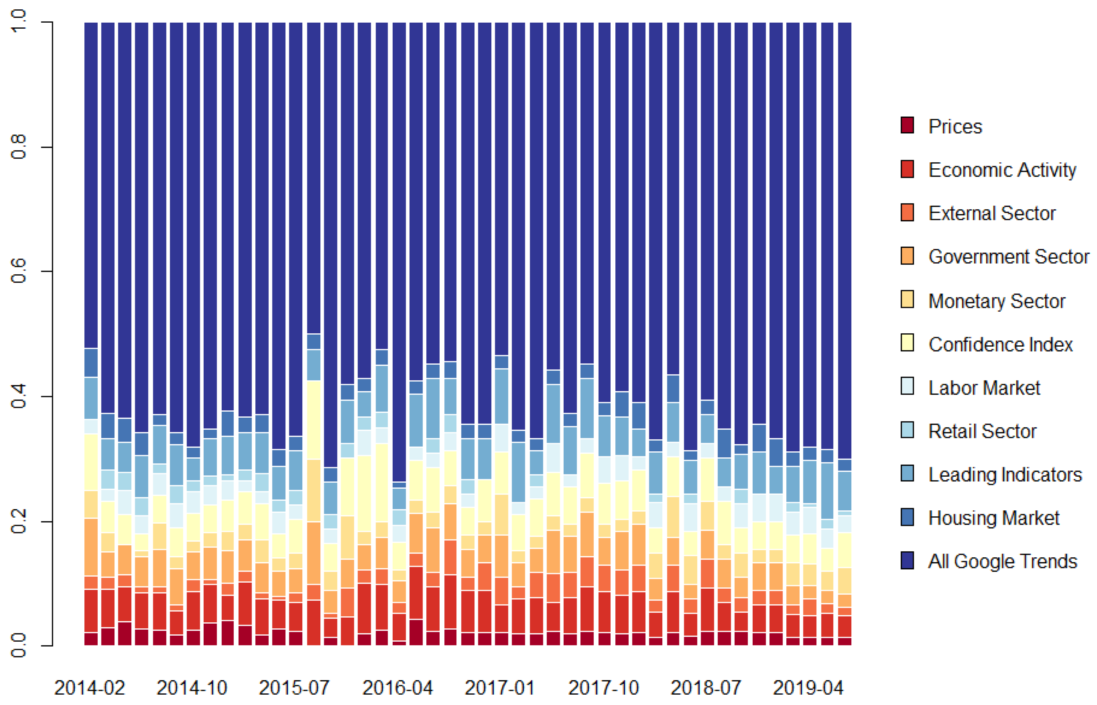


(d) Economic Indicators, US

Figure A.5: Variable Selection in MRP Approach (cont.)



(e) Economic Indicators and Google Trends Categories, United States



(f) Economic Indicators, Google Trends Categories and Subcategories, United States

Figure A.5: Variable Selection in MRP Approach (cont.)

Table A.1: Forecast Evaluation of Factor Models with Google Trends, 2014-2016

	Forecast ($h = 1$)			Nowcast ($h = 0$)			Backcast ($h = -1$)	
	1	2	3	1	2	3	1	2
Brazil								
Economic Indicators and Google Trends categories								
All	1.11	1.14	1.15	1.12	1.04	0.97	**0.79***	**0.77***
LASSO	**0.96**	**0.94**	**0.88***	**0.89***	**0.86***	0.87*	0.80*	0.81*
AdaLASSO	1.11	1.10	1.00	1.10	1.04	**0.84***	0.91	0.87
ENET	0.99	1.01	1.00	1.02	1.10	1.03	1.05	1.04
MRP	1.04	1.01	1.07	1.01	1.03	1.00	0.90	0.88
Economic Indicators, Google Trends categories and subcategories								
All	1.37	1.42	1.30	1.41	1.37	1.33	1.35	1.39
LASSO	1.29	1.29	1.31	1.29	1.28	1.31	1.30	1.26
AdaLASSO	1.16	1.20	1.15	1.19	1.14	0.93	0.91	0.82*
ENET	**1.03**	**1.13**	**1.09**	**0.95**	**0.97**	**0.83***	**0.84***	**0.80***
MRP	1.08	1.23	1.19	1.13	1.60	0.92	0.95	1.02
United States								
Economic Indicators and Google Trends categories								
All	0.97	1.02	0.99	1.00	**0.83***	**0.87**	0.88	
LASSO	**0.92**	**0.90**	**0.94**	**0.92**	0.95	0.96	0.97	
AdaLASSO	1.03	1.03	0.94	0.96	0.98	1.01	1.06	
ENET	1.00	1.00	0.97	0.95	0.96	0.98	0.97	
MRP	0.98	0.97	0.95	0.96	0.93	0.95	**0.86***	
Economic Indicators, Google Trends categories and subcategories								
All	**0.95**	**0.94**	**0.98**	1.14	**0.96**	1.05	1.05	
LASSO	1.02	1.06	1.11	1.08	1.06	1.05	1.06	
AdaLASSO	1.12	1.09	1.12	1.10	1.11	1.07	1.09	
ENET	1.06	1.03	1.08	1.07	1.04	1.05	1.04	
MRP	1.04	0.97	1.05	**1.03**	1.11	**1.01**	**0.98**	

Notes: Entries in this table denote relative Root Mean Squared Forecast Errors (RMSFEs) of factor models constructed with Economic Indicators and Google search series compared to factor models based only on Economic Indicators. Therefore, entries lower than one suggest that a particular DFM specification that utilises Google data next to economic indicators is more accurate compared to factor models based on economic indicators of the same variable selection method, for a given forecast horizon. Numbers in bold indicate model specifications with the "locally–best" RMSFEs, for a given forecast horizon. Grey boxes denote the lowest error measure achieved by the competing models for a given country and forecast horizon. Numbers followed by asterisks (**5% level, *10% level) are significantly superior to the DFM based on only economic indicators according to the Diebold–Mariano test.

Table A.2: Forecast Evaluation of Factor Models with Google Trends, 2017-2019

	Forecast ($h = 1$)			Nowcast ($h = 0$)			Backcast ($h = -1$)	
	1	2	3	1	2	3	1	2
Brazil								
Economic Indicators and Google Trends categories								
All	1.17	1.16	1.11	1.10	0.95	**0.75***	**0.78***	0.81*
LASSO	1.08	**0.93**	**0.83**	**0.86***	**0.82***	0.92	0.87	0.85
AdaLASSO	1.08	1.11	1.03	1.09	1.04	0.96	0.93	0.93
ENET	1.10	1.00	1.15	1.14	0.97	1.12	1.06	1.05
MRP	**1.05**	1.08	1.02	1.10	1.02	0.92	0.81*	**0.79***
Economic Indicators, Google Trends categories and subcategories								
All	1.13	1.11	1.09	1.13	1.10	1.09	0.98	1.05
LASSO	**1.07**	1.10	1.13	1.08	1.10	1.06	1.02	1.01
AdaLASSO	1.27	1.12	1.32	1.48	1.00	1.26	1.39	1.33
ENET	1.07	**1.05**	**1.03**	**0.98**	**0.85***	**0.83***	**0.84**	**0.78***
MRP	1.28	1.08	1.22	1.17	1.30	1.22	1.11	0.99
United States								
Economic Indicators and Google Trends categories								
All	0.97	1.02	0.96	0.99	**0.87***	**0.86***	**0.89**	
LASSO	1.08	1.03	0.88*	**0.89***	0.92	0.99	1.03	
AdaLASSO	1.01	0.98	0.96	0.97	0.93	0.96	0.98	
ENET	1.02	**0.83***	0.95	0.96	0.97	1.03	0.94	
MRP	**0.95**	0.95	**0.86***	0.91	0.87*	0.89	1.09	
Economic Indicators, Google Trends categories and subcategories								
All	0.98	1.00	0.97	0.96	**0.93**	0.97	1.02	
LASSO	1.07	1.00	0.93*	0.94	0.96	1.01	1.00	
AdaLASSO	0.99	**0.91**	0.95	1.01	0.99	1.00	1.01	
ENET	1.09	1.04	**0.90***	**0.92**	0.99	1.01	**0.98**	
MRP	**0.91**	1.05	0.99	0.92	0.95	**0.93**	1.00	

Notes: Entries in this table denote relative Root Mean Squared Forecast Errors (RMSFEs) of factor models constructed with Economic Indicators and Google search series compared to factor models based only on Economic Indicators. Therefore, entries lower than one suggest that a particular DFM specification that utilises Google data next to economic indicators is more accurate compared to factor models based on economic indicators of the same variable selection method, for a given forecast horizon. Numbers in bold indicate model specifications with the "locally–best" RMSFEs, for a given forecast horizon. Grey boxes denote the lowest error measure achieved by the competing models for a given country and forecast horizon. Numbers followed by asterisks (**5% level, *10% level) are significantly superior to the DFM based on only economic indicators according to the Diebold–Mariano test.

Table A.3: Forecast Evaluation of Factor Models (IC_{p1})

	Forecast ($h = 1$)			Nowcast ($h = 0$)			Backcast ($h = -1$)	
	1	2	3	1	2	3	1	2
Brazil								
Economic Indicators								
All	1.01	0.90	0.88*	**0.84***	**0.80****	0.62**	0.74*	0.71*
LASSO	**0.99**	0.97	**0.88****	0.93	0.87	0.64**	0.67*	0.64*
AdaLASSO	1.14	1.10	0.98	1.01	0.98	0.77**	0.79*	0.80*
ENET	1.22	1.11	0.98	0.98	0.85*	0.65**	0.64*	0.62*
MRP	1.01	**0.89***	0.99	0.96	0.74**	**0.58****	**0.59***	**0.55***
Economic Indicators and Google Trends categories								
All	1.22	1.18	1.04	1.07	0.78**	**0.50****	**0.56***	**0.57***
LASSO	**1.06**	**0.86**	**0.83***	**0.90**	**0.73***	0.63**	0.67*	0.64*
AdaLASSO	1.33	1.24	1.03	1.15	1.02	0.68**	0.72*	0.72*
ENET	1.22	1.09	1.01	1.02	0.89*	0.70**	0.68*	0.65*
MRP	1.14	1.08	0.95	0.91*	0.89*	0.62**	0.65*	0.63*
Economic Indicators, Google Trends categories and subcategories								
All	1.89	1.87	1.82	1.85	1.63	1.30	1.35	1.31
LASSO	1.21	1.22	1.10	1.13	1.09	0.83*	0.87*	0.87
AdaLASSO	1.34	1.28	1.15	1.27	1.07	0.82*	0.88	0.83*
ENET	1.23	1.14	**1.04**	0.95	**0.77****	0.57**	**0.53***	**0.50***
MRP	**1.16**	**1.02**	1.10	**0.94**	0.87	**0.55****	0.57*	0.55*
United States								
Economic Indicators								
All	0.85	0.91*	0.91*	**0.78***	0.85	0.75	**0.74**	
LASSO	**0.82**	**0.90**	**0.85***	0.80*	0.82	**0.75***	0.75	
AdaLASSO	0.84	0.90	0.94	0.91	0.86	0.83	0.81	
ENET	0.83	0.91	0.97	0.92	0.81*	0.76*	0.76	
MRP	0.83	0.90*	0.89*	0.82	**0.78***	0.80	0.79	
Economic Indicators and Google Trends categories								
All	0.84	0.93	0.91*	**0.80***	0.83	0.75	0.74	
LASSO	0.84	0.87	**0.83****	0.83	0.82	**0.73***	0.75	
AdaLASSO	0.81	0.87	0.89*	0.87	0.81*	0.77*	0.77	
ENET	**0.80**	**0.86**	0.89*	0.83	0.82	0.76*	0.73	
MRP	0.83	0.93	0.90	0.81*	**0.78***	0.74*	**0.71***	
Economic Indicators, Google Trends categories and subcategories								
All	**0.81**	0.91	**0.83***	0.87	**0.81**	**0.77***	0.79	
LASSO	0.87	0.95	0.98	0.87	0.88	0.78*	**0.77**	
AdaLASSO	0.90	0.91	0.89*	0.87	0.83*	0.82*	0.86	
ENET	0.87	0.94	0.90	**0.82***	0.83*	0.78*	0.80	
MRP	0.84	**0.88**	0.87*	0.85	0.84	0.83	0.82	

Notes: Entries in this table show Root Mean Squared Forecast Errors (RMSFEs) of Dynamic Factor Models (DFMs) relative to the AR(1) benchmark. Therefore, entries lower than one suggest that a particular prediction based on a DFM is more accurate compared to the AR(1) benchmark. Numbers in bold indicate predictions with the lowest relative RMSFE for each forecast horizon, within a given DFM specification. Grey boxes denote the lowest error measure achieved by the competing models for a given country and forecast horizon. Numbers followed by asterisks (**5% level, *10% level) are significantly superior to the AR(1) benchmark based on the Diebold–Mariano test.

Table A.4: Data Description – Brazil

The table below lists the *Key Economic Indicators* derived from Bloomberg. The fifth column reports the approximate delays, in days, with which indicators are released. This is necessary in order to construct the *pseudo* real–time dataset and to make sure that we never use data that would have not been available at that time. The final column shows the transformations employed to achieve stationarity. Zero indicates no transformation, two represents a monthly difference, four monthly difference in year over year difference, five year over year difference and six percentage change in year over year. Note that all economic indicators are on monthly frequency. Table A.4 contains Brazil's dataset and table A.5 contains United States' dataset.

No.	Mnemonic	Description	Group	Lag	Transf.
1	BZIPTL%	Industrial Production	Economic Activity	41	0
2	BZVPTLVH	Vehicle Production	Economic Activity	11	4
3	BZIPTLSA	Industrial Production Index	Economic Activity	41	2
4	BZCNCNIS	Manufacture Industry Capacity Utilization	Economic Activity	46	2
5	BZEAMOM%	Economic Activity GDP MoM%	Economic Activity	48	5
6	BZEAYOY%	Economic Activity GDP YoY%	Economic Activity	48	0
7	BZIPYOY%	Industrial Production	Economic Activity	41	0
8	BZCNCNI	Manufacture Industry Capacity Utilization	Economic Activity	46	5
9	BZCNEMPS	Manufacture Industry Employment	Economic Activity	46	2
10	BZCNSALS	Manufacture Industry Real Sales	Economic Activity	46	2
11	BZCNHOUS	Manufacture Industry Working Hours	Economic Activity	46	2
12	BZCACURR	Current Account Monthly	External Sector	28	4
13	BZTBBALM	Total Trade Balance FOB (USD mn)	External Sector	3	4
14	BZFDTMON	Foreign Direct Investment Net	External Sector	28	4
15	BZTBEXPM	Exports FOB (USD mn)	External Sector	3	4
16	BZTBIMPM	Imports FOB (USD mn)	External Sector	3	4
17	BZCA%GDP	Current Account (% of GDP)	External Sector	28	5
18	BZVXEXTL	Vehicle Exports	External Sector	11	4
19	BZBPCAPT	BOP Capital Account Net (USD mn)	External Sector	28	4
20	BZBPFINA	BOP Financial Account Net (USD mn)	External Sector	28	4
21	BZBPCURA	BOP Current Account Net (USD mn)	External Sector	28	4
22	BZDPNDT%	Public Net Debt % of GDP	Government Sector	31	2
23	BZBGPRIM	Central Government Primary Budget Balance (BRL bn)	Government Sector	30	4
24	BSRFTOFD	Total Federal Revenue (BRL mn)	Government Sector	24	4
25	BZPBPRDM	Primary Budget Result (BRL bn)	Government Sector	31	5
26	BZPBNODM	Nominal Budget Result (BRL bn)	Government Sector	31	4
27	BZPBNO%	Nominal Budget Result % of GDP 12 Month Flows	Government Sector	46	5
28	BZPBPR%	Primary Budget Result % of GDP 12 Month Flows	Government Sector	57	5
29	BZBGNOMI	Central Government Nominal Budget Balance (BRL bn)	Government Sector	30	6
30	BSRFTFY	Total Federal Revenue YoY%	Government Sector	24	0
31	BSRFTOMM	Total Federal Revenue MoM%	Government Sector	24	5
32	IBRENCMM	Construction Prices INCC-M MoM %	Housing Market	-2	5
33	IBRENCMY	Construction Prices INCC-M YoY%	Housing Market	-1	2
34	BZMW	Minimum Wage (BRL)	Labor Market	0	6
35	BZJCGTOT	Government Registered Job Creation Total	Labor Market	28	4
36	OEBRKLAR	OECD Leading Indicators CLI Trend Restored	Leading Indicators	7	2
37	OEBRKLAP	OECD Leading Indicators CLI Trend Restored YoY%	Leading Indicators	7	2
38	BZLNTOTA	Financial System Loans	Monetary Sector	29	4
39	BRCDDEFT	Personal Loans More Than 90 Days Late (% of total loans)	Monetary Sector	29	5
40	BZLNTMOM	Financial Total Outstanding Loans MoM%	Monetary Sector	29	5
41	BZMS1Y%	Money Supply M1 YoY%	Monetary Sector	29	0

Continued on next page

Table A.4 – *Continued from previous page*

No.	Mnemonic	Description	Group	Lag	Transf.
42	BZMS2Y%	Money Supply M2 YoY%	Monetary Sector	29	2
43	BZMBMB	Monetary Base (BRL bn)	Monetary Sector	29	4
44	BZMS2	Money Supply M2 Brazil M2	Monetary Sector	29	4
45	BZMS1	Money Supply M1 Brazil M1 (BRL bn)	Monetary Sector	29	4
46	BZMS3	Money Supply M3	Monetary Sector	29	4
47	BZMS1M%	Money Supply M1 MoM%	Monetary Sector	29	5
48	BZMS3Y%	Money Supply M3 YoY%	Monetary Sector	29	2
49	BZMS2M%	Money Supply M2 MoM%	Monetary Sector	29	5
50	BZMS4Y%	Money Supply M4 YoY%	Monetary Sector	29	2
51	BZMS3M%	Money Supply M3 MoM%	Monetary Sector	29	5
52	BZMS4M%	Money Supply M4 MoM%	Monetary Sector	29	5
53	BZMS4	Money Supply M4 Brazil M4	Monetary Sector	29	4
54	BRCDDESH	Personal Loans from 15 to 90 Days Late (% of total loans)	Monetary Sector	35	5
55	BZMBMON%	Monetary Base MoM%	Monetary Sector	30	5
56	BZMBYOY%	Monetary Base YoY%	Monetary Sector	21	0
57	BZPIIPCY	CPI IPCA YoY%	Prices	12	2
58	BZPIIPCM	CPI IPCA MoM%	Prices	12	5
59	BZPIIPMO	IPCA-15 CPI Extended National MoM%	Prices	-4	5
60	IBREGPMM	CPI IGP-M MoM%	Prices	-1	5
61	IBREGPMY	CPI IGP-M YoY%	Prices	-1	2
62	IBREGPDM	CPI IGP-DI MoM%	Prices	9	5
63	BZPIIPYO	IPCA-15 CPI Extended National YoY%	Prices	-4	2
64	IBREGPDY	CPI IGP-DI YoY%	Prices	9	2
65	BIGPIGP1	CPI First 10 Day Period Preview	Prices	-18	5
66	BZCPI	CPI FIPE MoM%	Prices	7	5
67	IBREGP1M	CPI IGP-10 MoM%	Prices	-14	5
68	BZPIINPM	CPI INPC MoM%	Prices	11	5
69	BIGPIGP2	CPI Second 10 Day Period Preview	Prices	-10	5
70	BZPIINPY	CPI INPC YoY%	Prices	12	0
71	IBREPCDM	CPI IPC-DI MoM%	Prices	-2	5
72	IBREPAMY	Wholesale Prices IPA-M YoY%	Prices	-1	2
73	IBREPAMM	Wholesale Prices IPA-M MoM%	Prices	-1	5
74	IBREPADY	Wholesale Prices IPA-DI YoY%	Prices	9	2
75	IBREPCDY	CPI IPC-DI YoY%	Prices	9	2
76	IBREGP1Y	CPI IGP-10 YoY	Prices	-14	2
77	IBREPCMY	CPI IPC-M YoY%	Prices	-1	2
78	IBREPCMM	CPI IPC-M MoM%	Prices	-1	5
79	IBREPADM	FGV Brazil Wholesale Prices IPA-DI MoM%	Prices	9	5
80	IBREPC1M	CPI IPC-10 MoM%	Prices	-14	5
81	IBREPC1Y	CPI IPC-10 YoY%	Prices	-14	2
82	BZRTRYOY	Retail Sales Volume Monthly YoY	Retail Sector	46	0
83	BZRTRETM	Retail Sales Volume MoM% Change	Retail Sector	46	5
84	BZVLTLVH	Vehicle Sales Licensed (total licensed vehicles)	Retail Sector	11	4
85	BZVLTOTL	Vehicle Sales Licensed Cars (total licensed cars)	Retail Sector	11	4
86	BZRTRIYY	Retail Sales Revenue	Retail Sector	46	0
87	BZRTCYOY	Retail Sales Volume Clothes & Footwear	Retail Sector	46	0
88	BZRTFUYY	Retail Sales Volume Furniture & Domestic Appliance	Retail Sector	46	0
89	BZRTFYYO	Retail Sales Volume Supermarket & Food & Beverage & Tobacco	Retail Sector	46	0
90	BZRTSYOY	Retail Sales Volume Supermarkets	Retail Sector	46	0
91	BZRTFYOY	Retail Sales Volume Fuel & Lubricants	Retail Sector	46	0
92	BZASSUBT	Auto Sales Subtotal	Retail Sector	5	4
93	BZRTAMPM	Amplified Retail Sales Volume MoM	Retail Sector	46	0
94	BZRTAMPY	Amplified Retail Sales Volume YoY	Retail Sector	46	0

Continued on next page

Table A.4 – *Continued from previous page*

No.	Mnemonic	Description	Group	Lag	Transf.
95	BZICINDX	Industrial Confidence General	Survey Indicators	-7	4
96	BZCCI	Consumer Confidence	Survey Indicators	5	5

Table A.5: Data Description – United States

No.	Mnemonic	Description	Group	Lag	Transf.
1	CPTICHNG	Capacity Utilization % of Total Capacity	Economic Activity	18	2
2	DGNOCHNG	Durable Goods New Orders Industries MoM	Economic Activity	66	0
3	DGNOYOY	Durable Goods New Orders Total YoY	Economic Activity	66	0
4	IP CHNG	Industrial Production MoM	Economic Activity	18	0
5	IP YOY	Industrial Production YoY	Economic Activity	18	0
6	MGT2MA	MFG+TRD INV/SALES RATIO MANUFACT	Economic Activity	76	2
7	MGT2RE	MFG+TRD INV/SALES RATIO RETAIL	Economic Activity	76	2
8	MGT2TB	MFG+TRD INV/SALES RATIO TOTAL BUSI	Economic Activity	76	2
9	MGT2WH	MFG+TRD INV/SALES RATIO WHOLESALE	Economic Activity	76	2
10	MGT2WHDU	MFG+TRD INV/SALES RATIO WHOLE DURA	Economic Activity	76	2
11	MTIBCHNG	Manufacturing & Trade Inventories Total MoM	Economic Activity	76	0
12	MWINCHNG	Merchant Wholesalers Inventories Total Monthly % Change	Economic Activity	62	0
13	TMNOCHNG	Manufacturers New Orders Total MoM	Economic Activity	66	0
14	FRNTTNET	Treasury International Capital Net Monthly Inflows Total	External Sector	62	5
15	FRNTTOTL	Foreign Net Transactions	External Sector	62	5
16	IMP1CHNG	Import Price Index by End Use All MoM NSA	External Sector	16	5
17	IMP1YOY%	Import Price Index by End Use All YoY NSA	External Sector	16	2
18	USTBEXP	Trade Balance of Exports	External Sector	66	2
19	USTBEXPY	Trade Balance of Exports YoY	External Sector	66	2
20	USTBIMP	Trade Balance of Imports	External Sector	66	2
21	USTBIMPY	Trade Balance of Imports YoY	External Sector	66	2
22	USTBTOT	Trade Balance of Goods and Services SA	External Sector	68	2
23	DEBPBILL	Total Debt Outstanding Bills	Government Sector	44	4
24	DEBPBOND	Total Debt Outstanding Bonds	Government Sector	44	4
25	DEBPINNT	Total Debt Outstanding TIPS	Government Sector	44	4
26	DEBPMARK	Total Debt Outstanding Total Marketable	Government Sector	44	4
27	DEBPNMRK	Total Debt Outstanding Total Non Marketable	Government Sector	44	4
28	DEBPNOTE	Total Debt Outstanding Notes	Government Sector	44	4
29	DEBPTOTL	Total Debt Outstanding	Government Sector	44	4
30	FDDSGDP	Treasury Fed. Budget Deficit/Surplus as %Nominal GDP	Government Sector	44	4
31	FDDSSD	Treasury Fed. Budget Debt Summary Deficit/Surplus	Government Sector	44	5
32	ETSLMOM	Existing Homes Sales MoM	Housing Market	23	0
33	ETSLTOTL	Existing Homes Sales	Housing Market	23	2
34	HPIMMOM%	FHFA US House Price Index Purchase Only MoM%	Housing Market	61	0
35	MBAVCHNG	MBA Mortgage Market Index Weekly % Change	Housing Market	5	0
36	NHSLTOT	New One Family Houses Sold Annual Total	Housing Market	64	2
37	NHSPATOT	Priv. Housing Authorized by Bldg Permits by Type Total	Housing Market	57	2
38	NHSPSTOT	New Priv. Owned Housing Units Started by Structure Total	Housing Market	57	2
39	SPCS20Y%	Case-Shiller 20-City Comp. City Home Price Index YOY%	Housing Market	61	2
40	USHBMIDX	National Association of Home Builders Market Index	Housing Market	-9	2
41	USPHTMOM	Pending Home Sales Index MoM	Housing Market	0	0
42	ADP CHNG	ADP National Employment Report Priv. Nonfarm Level Change	Labor Market	6	0
43	CHALYOY%	Challenger US Job Cut Announcements YoY% Change	Labor Market	7	0
44	EMDINP1M	Employment Diffusion Nonfarm Payrolls +1 Month	Labor Market	8	0
45	INJCJC	Initial Jobless Claims	Labor Market	12	2
46	INJCSP	Continuing Jobless Claims	Labor Market	12	2
47	JOLTHIRS	Hires Rate	Labor Market	47	2
48	JOLTOPEN	Job Openings Rate	Labor Market	47	2
49	JOLTSEPS	Separations Rate	Labor Market	47	2
50	NFP PCH	Employees on Nonfarm Payrolls Total Private MoM Net Change	Labor Market	8	0

Continued on next page

Table A.5 – *Continued from previous page*

No.	Mnemonic	Description	Group	Lag	Transf.
51	NFP TCH	Employees on Nonfarm Payrolls Total MoM Net Change	Labor Market	8	2
52	USEMNCHG	Employment Total in Labor Force Net Change	Labor Market	8	0
53	USEMTOT	Employment Total in Labor Force SA	Labor Market	8	2
54	USERTOT	Employment Population Ratio Total in Labor Force	Labor Market	8	2
55	USHETOT%	Avg Hourly Earnings Priv. Nonfarm Payrolls Total Nominal MoM	Labor Market	8	0
56	USHEYOY	Avg Hourly Earnings Priv. NFP Prod&NonSup in Nom$ YoY	Labor Market	8	2
57	USMMMNCH	Employees Nonfarm Payrolls Manuf. Industry Monthly Net Chg	Labor Market	8	0
58	USUDMAER	Unemployed & Part Time & Margin % Labor Force & Margin	Labor Market	8	2
59	USURTOT	Unemployment Rate Total in Labor Force	Labor Market	8	2
60	USWHTOT	Avg Wekkly Hours Nonfarm Total Priv. Production & Nonsupervisory	Labor Market	8	2
61	IP	Industrial Production	Leading Indicators	18	2
62	LEI ACE	Leading Index Average Consumer Expectation	Leading Indicators	24	0
63	LEI AVGW	Leading Index Avg Workweek Production Workers Manuf Hours	Leading Indicators	24	2
64	LEI BP	Leading Index Building Permits	Leading Indicators	24	2
65	LEI CHNG	Leading Index MoM	Leading Indicators	24	0
66	LEI IRTE	Leading Index Interest Spread 10 year Treasury Less Fed Fu	Leading Indicators	24	2
67	LEI LCI	Leading Index Leading Credit Index	Leading Indicators	24	0
68	LEI MNO	Manufacturers New Orders Nondefense Capital Good Ex Aircra	Leading Indicators	24	2
69	LEI NWCN	Leading Index Manuf New Orders Consumer Goods & Materials	Leading Indicators	24	2
70	LEI STKP	Leading Index Stock Prices 500 Common Stocks	Leading Indicators	24	2
71	LEI TOTL	Leading Index Ten Economic Indicators	Leading Indicators	24	2
72	LEI WKIJ	Leading Index Avg Weekly Initial Jobless Claims	Leading Indicators	24	2
73	LEI YOY	Leading Index Ten Economic Indicators YoY	Leading Indicators	24	2
74	MTSLRL$	Manufacturing & Trade Sales in Millions of chained (2009) dollars	Leading Indicators	46	2
75	NAPMNEWO	ISM Manufacturing Report on Business New Orders	Leading Indicators	3	0
76	NFP T	Employees on Nonfarm Payrolls Total	Leading Indicators	8	2
77	OEUSKLAP	OECD Leading Indicators CLI Trend Restored YoY	Leading Indicators	62	2
78	OEUSKLARI	OECD Leading Indicators CLI Trend Restored	Leading Indicators	62	2
79	PIDSPINX	Personal Income Excl Transfer Receipts Chained 2012 Dollars	Leading Indicators	59	2
80	ARDIMONY	Monetary Base Total	Monetary Sector	8	4
81	ARDIMOYY	Monetary Base Total YoY	Monetary Sector	8	2
82	CICRTOT	Federal Reserve Consumer Credit Total Net Change	Monetary Sector	40	0
83	M1% YOY	Federal Reserve Money Supply M1 YoY % Change	Monetary Sector	23	2
84	M2% YOY	Federal Reserve Money Supply M2 YoY % Change	Monetary Sector	23	2
85	PCE CHNC	Personal Consumption Expenditure Chained 2012 Dollars MoM	Monetary Sector	59	0
86	PCE CHY%	Personal Consumption Expenditure Chained 2012 Dollars YoY	Monetary Sector	59	2
87	PIDSDPS	Personal Saving as a % of Disposable Personal Income	Monetary Sector	59	2
88	PITLCHNG	Personal Income MoM	Monetary Sector	59	0
89	PITLYOY	Personal Income YoY	Monetary Sector	59	0
90	CPI CHNG	CPI Urban Consumers MoM	Prices	13	0
91	CPI XYOY	CPI Urban Consumers Less Food & Energy YoY	Prices	13	0
92	CPI YOY	CPI Urban Consumers YoY	Prices	13	0
93	CPUPXCHG	CPI Urban Consumers Less Food & Energy MoM	Prices	13	0
94	PCE CYOY	Personal Consumption Expenditure Core Price Index YoY	Prices	59	2
95	PCE DEFY	Personal Consumption Expenditures Chain Type Price Index YoY	Prices	59	2
96	RSTAMOMx	Adjusted Retail & Food Services Sales SA Total Monthly % Change	Retail Sector	35	0
97	RSTAXMOM	Adjusted Retail Sales Less Autos SA Monthly % Change	Retail Sector	35	0
98	RSTAXYOY	Adjusted Retail Sales Less Autos Yearly % Change	Retail Sector	45	2
99	RSTAYOY	Adjusted Retail & Food Services Sales Total Yearly % Change	Retail Sector	45	2
100	SAARDTOT	Auto Sales Domestic Vehicles Annualized	Retail Sector	4	2
101	SAARTOTL	Auto Sales Total Annualized	Retail Sector	4	2
102	CFNAI	Chicago Fed National Activity Index	Survey Indicators	28	5
103	CFNAIMA3	Chicago Fed National Activity Index Three Month Mov. Av.	Survey Indicators	28	5

Continued on next page

Table A.5 – *Continued from previous page*

No.	Mnemonic	Description	Group	Lag	Transf.
104	CHPMINDX	Market News International Chicago Business Barometer	Survey Indicators	0	0
105	COMFBTWR	Bloomberg US National Economy Expectations Diffusion Index	Survey Indicators	-9	4
106	COMFCOMF	Bloomberg US Weekly Consumer Comfort Index	Survey Indicators	0	5
107	CONCCONF	Conference Board Consumer Confidence	Survey Indicators	0	2
108	CONSSENT	University of Michigan Consumer Sentiment Index	Survey Indicators	0	5
109	EMPRGBCI	Manufacturing Survey General Business Conditions	Survey Indicators	-13	0
110	MAPMINDX	ISM Milwaukee Purchasers Manufacturing Index	Survey Indicators	0	0
111	NAPMNMI	ISM Services PMI	Survey Indicators	7	4
112	NAPMPMI	ISM Manufacturing PMI	Survey Indicators	3	2
113	NAPMPRIC	ISM Manufacturing Report on Business Prices Index	Survey Indicators	3	4
114	OUTFGAF	Phil. Fed Business Outlook Survey Diffusion Index General Conditions	Survey Indicators	-9	0
115	RCHSINDX	Richmond Manuf. Survey Current Manuf. Composite	Survey Indicators	-4	0

Chapter 3

The Predictive Power of Internet Search Data in Forecasting Private Consumption

3.1 Introduction

The continuous assessment of macroeconomic conditions is of primary importance. The Global Financial Crisis of 2007-2009 highlighted the necessity for real-time monitoring of economic and financial variables. The Covid-19 pandemic further demonstrated that when conditions in the real economy evolve rapidly, relying on traditional – backward-looking – economic indicators may not be an optimal approach from a forecasting perspective.

Macroeconomic indicators are typically available at low frequencies (e.g., monthly, quarterly), published with significant delays, and often are subject to substantial revisions. As a response to these limitations, alternative sources of high-frequency data have emerged, providing a continuous assessment of current economic developments. One prominent source of alternative data is Google Trends, a powerful – freely available – dataset that shows the relative popularity of search queries on the Google search engine.

A growing body of literature supports the idea of using internet search data for the prediction of key macroeconomic indicators, *inter alia*, Götz and Knetsch (2019), Woloszko (2020), Borup and Schütte (2022).

This chapter aims to evaluate the predictive ability of Google Trends data in forecasting U.S. private consumption and its components, that is, durable goods, nondurable goods, and services. Given that private consumption accounts for approximately 70% of the U.S. gross domestic product, it serves as a major driver of economic growth. Consequently, timely and accurate predictions of private consumption is of utmost importance for policymakers, businesses, and households.

Given the rapid shift of consumers to online shopping, it is reasonable to assume that latent consumer sentiment can be crystallised on internet search behaviour, particularly among individuals who are at the pre-purchase stage. While conventional macroeconomic indicators reflect the *ability* of individuals to spend and survey data convey their sentiment about the *willingness* to spend, Google search data could provide an additional dimension, capturing the *preparatory steps* involved in the purchasing process. Furthermore, the Covid-19 pandemic has sharply accelerated the shift towards digital consumption, as evidenced by a 2020 survey conducted by UNCTAD[1], indicating that this trend is likely to outlast the pandemic. Therefore, consumers continually leave a substantial search footprint, and leveraging this information can potentially lead to sizeable forecasting advantages.

We have constructed our consumption-related Google Trends dataset based on 22 primitive keywords that are closely aligned with the Bureau of Economic Analysis (BEA) classification of primary consumption components. Selecting the primitive Google Trends series using such a subjective way may provide more gains compared to data-driven selection methods (Combes and Bortoli, 2016). To expand the dataset, we utilised Google Trends' "related queries" function, resulting in a total of 129 series. In addition, we have created two alternative Google Trends datasets: Google Trends "categories" and "keyword planner". The Google Trends "categories" dataset follows a hierarchical classifica-

[1] For more information see https://unctad.org/press-material/covid-19-has-changed-online-shopping-forever-survey-shows

tion with five levels, where each category encompasses a broad range of related keywords. The categories are selected on a judgemental basis, following a similar procedure as in Vosen and Schmidt (2011) and Woo and Owen (2019). The Google Trends "keyword planner" dataset, on the other hand, leverages a free Google product designed mainly to assist marketers in keyword research, providing data on average monthly searches, competition, cost, and more. Constructing this dataset follows a similar procedure to "related queries", where we start with each primitive query and extract the first five relevant keywords with the highest average monthly search volumes.

Although Google Trends data suffer from some limitations such as sampling error, location shifts, and downward trend bias, they contain some appealing features over traditional macroeconomic series within a forecasting framework. First, the majority of macroeconomic variables are only available at lower frequencies, and are typically published with a significant delay. For example, in the United States, official data publications on private consumption are released on a monthly basis with a publication lag of around a month. In contrast, Google search data are available on a daily frequency and are published with a negligible delay. Second, several economic indicators are subject to substantial revisions. Third, although survey data are typically available in a timely fashion and are not revised, sometimes obtaining such data can be costly or they may not be available at all. On the other hand, a forecaster can easily extract Google Trends data in real-time, tailor the data to specific geographical areas, and download data for more than 180 countries.

Given the substantial expansion of both traditional and alternative indicators across temporal and cross-sectional dimensions, it is of paramount significance for economic forecasters to avoid conventional estimation techniques since their performance tends to wane as the dimensionality of explanatory variables rises. Typically, when dealing with large datasets, economists often employ sparse, dense, or machine learning modelling. Sparse models lead to a small set of regressors with the highest forecasting power (e.g., LASSO), dense modelling assumes that all series might be important for prediction (e.g., factor models), and machine learning models can accommodate for non-linearities (e.g., random forests). The common feature of these methods is the presence of dimensionality reduction that aims to deliver more accurate predictions and to restrict overfitting issues.

Thus, to assess the potential predictive power of Google search data, our methodological framework includes three alternative classes of models: shrinkage, factor, and ensemble & machine learning models. We then compare the performance of these models against three benchmarks: an autoregressive model, a linear regression model with a limited set of macroeconomic variables (Carroll et al., 1994; Ludvigson and Bram, 1998; Croushore, 2005), and a factor-augmented autoregressive model in which factors are computed from 128 monthly macroeconomic indicators from McCracken and Ng (2016). We benchmark our forecasts against the simple autoregressive model since it is considered to be the main starting point for short-term forecasting experiments, while we add the two macroeconomic benchmarks to assess the ability of Google Trends indicators to upgrade forecasting methods that already incorporate macroeconomic variables. This comprehensive benchmarking approach allows us to gauge the added value of Google Trends indicators in a rigorous and systematic way.

To provide a broader evaluation of the usefulness of Google Trends series, we remove any potential benefit that arise from the usage of these high-dimensional and machine learning models by comparing the performance of each model specification using only Google Trends data versus a large set of traditional macroeconomic predictors. This comparison helps us understand the unique contribution of Google Trends indicators. Furthermore, our assessment extends across two distinct sub-periods, targeting four different consumption variables, and utilising three diverse Google Trends datasets. This multifaceted evaluation ensures a robust analysis of the effectiveness of Google Trends and the selected models in various contexts.

Although there are several papers that use Google Trends data within a forecasting framework (Carrière-Swallow and Labbé (2013), Götz and Knetsch (2019), Yu et al. (2019), Woloszko (2020), Borup and Schütte (2022), Ferrara and Simoni (2023)), only a few focus on the usefulness of Google Trends data to forecast private consumption. Vosen and Schmidt (2011) provide evidence that Google Trends consists of a promising alternative source of information in forecasting consumer spending in the United States. Using unobserved factors generated from Google data, the authors show that Google Trends outperform the two most popular consumer confidence indicators, that is, the University of Michigan Consumer Sentiment Index and the Conference Board Consumer Confidence Index. In a similar setting, Woo and Owen (2019) utilise consumption-related Google indicators to forecast private consumption in the United States. The main difference with Vosen and Schmidt (2011) is that they treat Google data as complementary – and not as substitutes – to consumer confidence indicators. Empirical findings suggest that Google Trends provide forecast benefits over and above consumer confidence indicators and deliver significant information about pre-consumption trends.

We contribute to the fast-growing literature of economic forecasting with large – alternative – datasets by utilising Google Trends and several econometric and machine learning methods for predicting private consumption. While our starting point is closely associated with Vosen and Schmidt (2011) and Woo and Owen (2019), we depart from the existing literature in several respects. Our first contribution involves the expansion of the dataset. We go beyond the conventional use of the main Google Trends "categories" and incorporate the "related-queries" and "keyword planner" series, broadening the scope of our investigation into the potential predictive power of Google search data. Second, we meticulously address the limitations of Google Trends series, including sampling error, long-term downward bias, and abrupt shifts. Third, the analysis is also extended by including a broad range of statistical and econometrics models suitable to handle high-dimensional data structures. By disentangling the predictive power of different model types, we provide valuable insights for policymakers and market practitioners. Finally, although the methodological design of our approach is related to that of Medeiros et al. (2021), we include both data-poor and data-rich benchmarks and we provide a time-varying predictability dimension in our forecasting exercise.[2]

Our findings can be summarised as follows. First, including a large set of Google

[2]Here the term "data-poor" describes models in which only a few variables appear in the forecasting model, typically selected on a judgemental basis or using some form of variable selection. On the other hand, "data-rich" methods refer to estimation strategies that employs *all* information in a large dataset and not just an "optimal" subset of predictors.

Search series into the forecaster's information set, combined with the use of shrinkage and random forest models, yields considerable gains. Using the Google Trends "related queries" for forecasting the aggregate U.S. consumption expenditures improves forecast accuracy at $h = \{6, 9\}$ for a wide range of models, with improvements against a macroeconomic benchmark of 15-20%. Second, sizeable forecast gains are realised when forecasting the PCE components separately. In particular, Google-based models appear to be a valuable source of information for forecasting durable goods expenditures at horizons of $h = \{3, 6, 9\}$, resulting in gains of around 15-40% compared to the macroeconomic benchmark model. For nondurable expenditures, Google-based models demonstrate satisfactory predictions mainly at $h = \{1, 2\}$ with improvements up to 20% against the benchmark. For services expenditures, benefits arise mostly at 9-months ahead predictions, with noticeable gains of up to 30% compared to the benchmark.

Second, it is crucial to disentangle the relative contributions to the forecast gains of the Google Trends data and the models used. Thus, we compare the performance of each model specification that uses a large set of macroeconomic data against the same model with Google Trends data as predictors. Empirical findings verify the strong benefits that arise when we use Google Trends data for forecasting durable goods expenditures, especially at $h = \{3, 6, 9\}$ with gains up to 50%. In the case of nondurable goods, we obtain some statistically significant improvements over the macroeconomic series mostly at horizons $h = \{6, 9\}$, meaning that the significant gains that we documented in the first part of the analysis at short-horizon predictions arise mainly from the usage of sophisticated forecasting models against a simple benchmark rather than the Google Trends data. In the case of services and the aggregate consumption expenditures, forecast gains appear again mostly at horizons $h = \{6, 9\}$ with improvements of up to 20%, suggesting the usefulness of search data in long-horizon forecasts compared to macroeconomic information.

Third, we show that models that impose sparsity provide the highest forecast accuracy. By contrast, specifications that utilise a dense approach such as factor models perform quite poorly during the out-of-sample exercise. In particular, random forest, a non-linear machine learning model, produces in general the lowest forecast errors ranking at the top in 53 out of 120 forecast estimations. The performance of random forests varies depending on the set of predictors, with the best performance observed when using "categories". LASSO regression performs well across all Google Trends datasets, while targeted factors and complete subset regressions have disappointing performance. Fourth, an examination of model performance over time reveals that during the early phase of the pandemic, most models exhibit equally poor performance, which is a common trend in econometric models during crises and recoveries.

Finally, given that shrinkage and random forests exhibit the best forecasting performance overall, we attempt to identify the selected Google Trends series during the out-of-sample period. For durable goods consumption, the series differ pre and post-pandemic, with a shift towards computer and gaming-related keywords. In contrast, forecasting nondurable goods consumption frequently involves "Walgreens" and keywords related to gasoline, beverages, and clothing. When forecasting services consumption, "public health" is prominent at horizon $h = 1$, while "transportation" and electricity and gambling-related keywords gain importance at horizons $h = 6$ and $h = 9$. This dynamic variable selection pattern highlights the changing predictive ability of Google Trends data based on the business cycle, with more variables included pre-pandemic compared to the period after 2020.

The rest of this chapter is organised as follows: Section 3.2 describes the U.S. private consumption data and the construction of the three Google Trends datasets, and Section 3.3 discusses the methodological framework. Section 3.4 presents the forecasting setup and empirical findings, while Section 3.5 summarises the main empirical results of this chapter. Finally, Appendix B contains additional findings and robustness analyses.

3.2 Data Description

The United States Personal Consumption Expenditures (PCE) measure the goods and services purchased by households and nonprofit institutions serving households (NPISHs).

According to the Bureau of Economic Analysis, goods are categorised into durable and nondurable goods. Durable goods are those with an average lifespan of at least three years, while nondurable goods are defined as consumer items that can be either immediately consumed or have an average lifespan of up to three years, necessitating successive purchases. Examples of durable goods include motor vehicles and household furnishings, while nondurables include items like food, beverages, and clothing. Services are considered products that cannot be stored or inventoried; they are typically consumed immediately at the place where the purchase occurred.

Figure 3.1 shows that the three components of personal consumption exhibit heterogeneous features and react differently to economic events. For example, by examining the y-axis scale, one can observe that durable goods display significantly higher volatility compared to the other two components, with year-over-year growth rates reaching as high as 53.4% and a minimum of -19.5% during the Covid-19 period (Table 3.2). This volatility arises from the fact that expenditures on durable goods are highly procyclical. In a growing economy, consumers tend to have greater spending power, leading to an increase in durable goods consumption expenditures. Conversely, during times of elevated uncertainty, consumers typically postpone their purchasing plans for these goods.

On the other hand, nondurable goods demonstrate a noticeably more stable behaviour, as the majority of these goods represent necessary items that consumers are likely to purchase regardless of economic conditions. Expenditures on services show interesting characteristics, maintaining a stable behaviour up to 2020. However, after the initial outbreak of Covid-19, volatility rose dramatically, recording the largest decline among all components at -21% in April 2020. Therefore, these unique features of the private consumption components require detailed forecasting experiments as well as the examination of alternative sources of information, such as Google search data.

Google Trends provides data about the frequency that a specific keyword is searched in a given geographical area. To facilitate comparisons between queries easier, search terms are normalised by the total amount of Google searches in the selected region, on a specific period. Instead of providing the actual resulting number, Google produces an

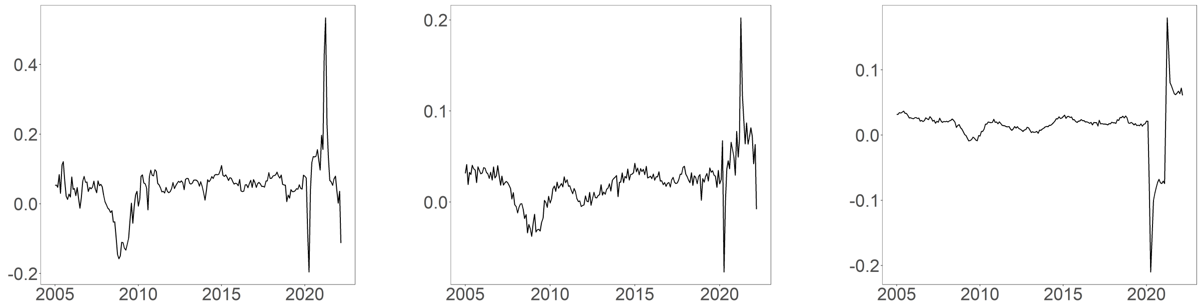


Figure 3.1: U.S. Private Consumption Expenditures Components

Notes: This graph shows the main components of U.S. private consumption expenditures on a year-over-year basis. The plot on the left-hand side shows durable goods consumption, the plot in the middle panel shows the nondurable goods consumption, while the plot on the right-hand side exhibits services consumption.

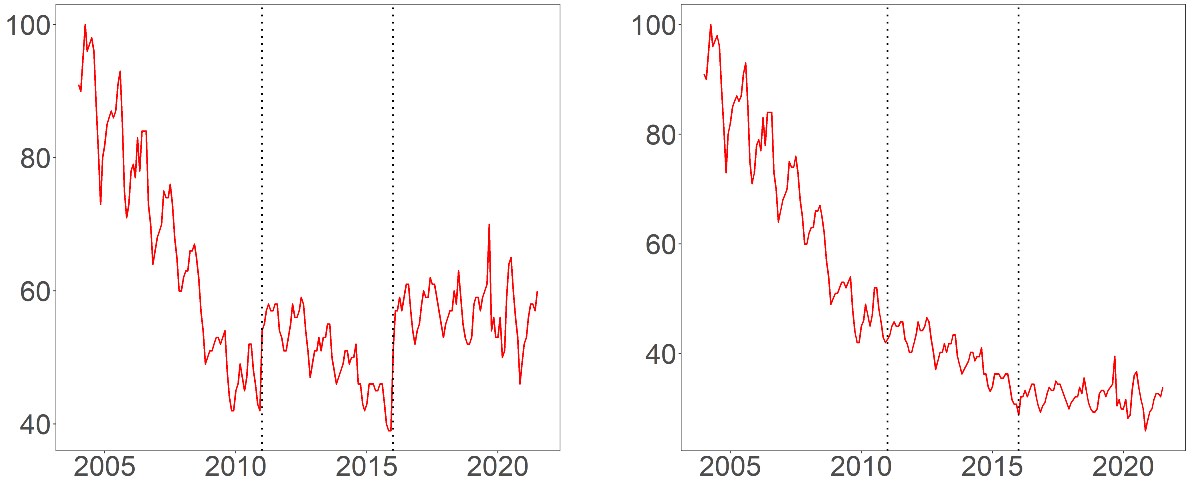
index that ranges between 0 and 100.

Although Google Trends data consists of a valuable source of information, it does not come without its drawbacks. First, Google Trends "categories" suffer from abrupt breaks in the series due to improvements in geographical assignments and data collection systems occurred in January 2011 and January 2016. Second, the series exhibit an underlying downward trend bias due to the constant increasing of the total search terms, that is, the denominator that normalises the actual search numbers.[3] Third, as a result of the way Google operates, the Google Trends data are subject to measurement error. This is because Google manages an incredibly large amount of queries on a daily basis and providing access to the entire set of data would be extremely time-consuming and inefficient. Thus, to construct the Google Trends series Google utilises only a small fraction of search data. As a result, if one extracts a specific query on two different days the series will not be the same.

We attempt to address each of these issues and appropriately clean the data before proceeding with the empirical analysis. There are sudden breaks in the Google Search series due to improvements in geographical assignments and data collection systems in January 2011 and January 2016 that need to be adjusted. As can be seen from Figure 3.2, which shows on the left-hand side the raw Google Trends "Autos & Vehicles" category,

[3]This feature is most prominent in the "categories" series. The "related queries" and the "keyword planner" series do not appear to suffer from this bias.

there are two abrupt changes in 2011 and 2016. The researcher needs to adjust for these breaks to avoid for potential distortions.



(a) Raw Series for "Autos & Vehicles" **(b)** Adjusted "Autos & Vehicles" for Breaks

Figure 3.2: Breaks in "Autos & Vehicles" Google Trends Category

Notes: There are two abrupt changes in Google Trends categories due to improvements in geographical assignments and data collection systems in January 2011 and January 2016 as shown by the two vertical dotted lines. The chart on the left-hand side shows the raw series for "Autos & Vehicles", while the chart on the right-hand side exhibits the adjusted series for breaks.

To solve this issue, a simple method is utilised in which post-break observations are multiplied by the ratio of the average of the three observations before the break over the average of the three post-break observations. The following formula is employed to address the break in the series in January 2011, and in the same way, we adjust for the break in 2016:

$$AdjustedSVI_{t=i} = SVI_{t=i} \times \left[\frac{N^{-1} \sum_j^n SVI_{t=j}}{N^{-1} \sum_z^n SVI_{t=z}} \right]$$

where $i = 85, \ldots, 208$, $j = 82, \ldots, 84$, $z = 85, \ldots, 87$ and $N = 3$. The two breaks occur in the 85th and 145th observation. The right panel of Figure 3.2, shows how we successfully removed the two breaks in the case of "Autos & Vehicles" data.

In order to understand the underlying trend bias issue we need first to understand what is a Google Trends series. Google Trends series essentially represent Search Volume

Indices (SVI) which can be constructed as follows: On a specific day t in a region r, for category c, the SVI is computed as the amount of queries for the particular category $Q_{t,r,c}$ divided by the total number of all search terms $TQ_{t,r}$. This ratio is then multiplied by a constant C_c to construct an index that ranges from 0 to 100:

$$SVI_{t,r,c} = \frac{Q_{t,r,c}}{TQ_{t,r}} \times C_c$$

The problem of the underlying downward trend bias arises because the denominator (total search terms) has surged substantially over the years, as more and more people use Google search engine more often, and as a result the number of total queries has increased. This downward common trend can be clearly observed from right-hand side of Figure 3.2, especially from 2004 to 2011.

In order to effectively remove this underlying common trend, we follow Woloszko (2020), and in which, first, the underlying trend of each series is extracted and then the first Principal Component is computed from all filtered long-term trends to reflect the common long-term trend component. Thereafter, the principal component is rescaled to have the same mean and standard deviation as the average of all Google series. Finally, the rescaled principal component is subtracted from the logarithmic SVIs

The *Hodrick-Prescott* filter, a popular filter among macroeconomists, is employed to extract the underlying trend of Google Trends series from short-run fluctuations. Specifically, the *HP filter* decomposes an observed series y_i into a trend component (τ_i) and a cyclical component (c_i), such that $y_i = \tau_i + c_i$. The trend component can be derived by solving the following constrained minimisation problem:

$$\min_{\tau_1,\ldots\tau_n} \sum_{i=1}^{n}(y_i - \tau_i)^2 + \lambda \sum_{i=2}^{n-1}(\tau_{i+1} - 2\tau_i + \tau_{i-1})^2$$

where λ is a non-negative smoothing parameter that penalises the variation in the long-term growth of the series. The larger the value of the smoothing parameter, the higher is the penalty. Ravn and Uhlig (2002) suggests that when dealing with monthly data the smoothing parameter should be equal to $129,600$.

The first principal component is extracted from the long-term filtered trends of Google Trends series and then is rescaled to have the same mean and standard deviation as the logarithmic Google Trends series.

$$RescaledPC = \overline{SVI} + (PC - \overline{PC}) \times \frac{\sigma_{SVI}}{\sigma_{PC}}$$

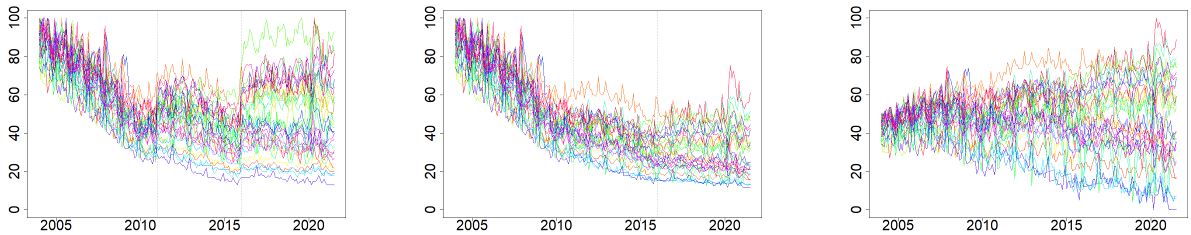
where $\overline{SVI}$ represents the mean value of all Google Trends series, $\overline{PC}$ denotes the mean value of the principal component, σ_{SVI} and σ_{PC} represent the standard deviations of SVIs and principal component. The right-hand plots in Figure 3.3 show the Google Trends series after the downward trend has been removed.

Another issue with Google Trends data is that Google's sampling methodology introduces a measurement error. This means that for a particular query on a certain day, Google Trends generates precisely the same series, but if the same keyword is extracted on different days, data series are not the same. The reason behind this sampling error is that Google is not able to provide access to the whole dataset of searches in a timely manner since it manages a tremendous amount of queries every day. Thus, instead of providing the actual search volumes, Google uses only a tiny sample of searches to construct the Google Trends series, and this particular sample is continuously changing.

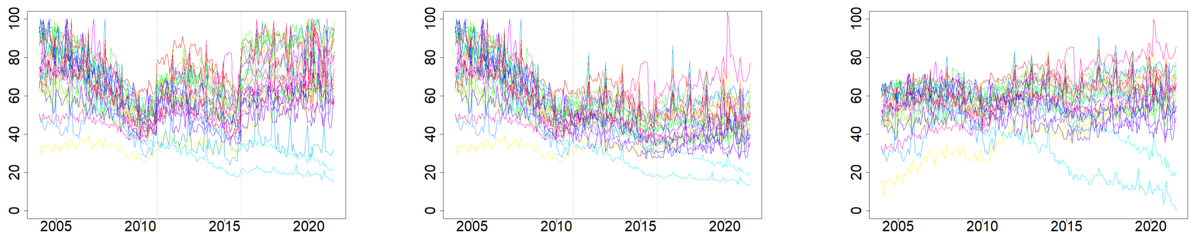
Figure 3.4 shows eight different samples of "Autos & Vehicle" Google Trends category, obtained in eight consecutive days for the same region and timestamp. As can be seen, it is clear that Google Trends data suffer from measurement error since each sample is marginally different from the other.

To deal with the sampling error, we follow Medeiros and Pires (2021) and compute the average of eight different Google Trends samples extracted on eight different days.[4] Figure 3.4 demonstrate the magnitude of the measurement error in the data while Table 3.1 indicates that by taking the simple average over multiple samples creates more reliable and stable series. Finally, it is worth noting that the literature has remained inconclusive

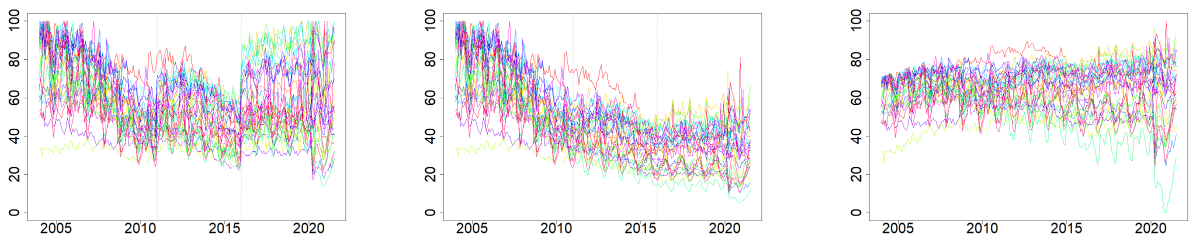
[4] Medeiros and Pires (2021) provide a detailed discussion about the fact that each Google Trends series is based on a subsample that is changing all the time and how this feature could be problematic in a forecasting framework. Using simulated and real data they explain that taking averages of multiple Google Trends samples improves forecast accuracy.



(a) Google Trends "Durable" Categories



(b) Google Trends "Nondurable" Categories



(c) Google Trends "Services" Categories

Figure 3.3: Google Trends Data Transformation

Notes: This graph shows the problem of abrupt breaks and downward trend bias in Google Trends data. Graphs on the left-hand side show the raw series for durable, nondurable and services Google Trends categories. Graphs in the middle show the Google Trends series after we have taken into consideration the two breaks occurred in 2011 and 2016. Finally, graphs on the right-hand side present the Google data after the downward long-term trend has been effectively removed.

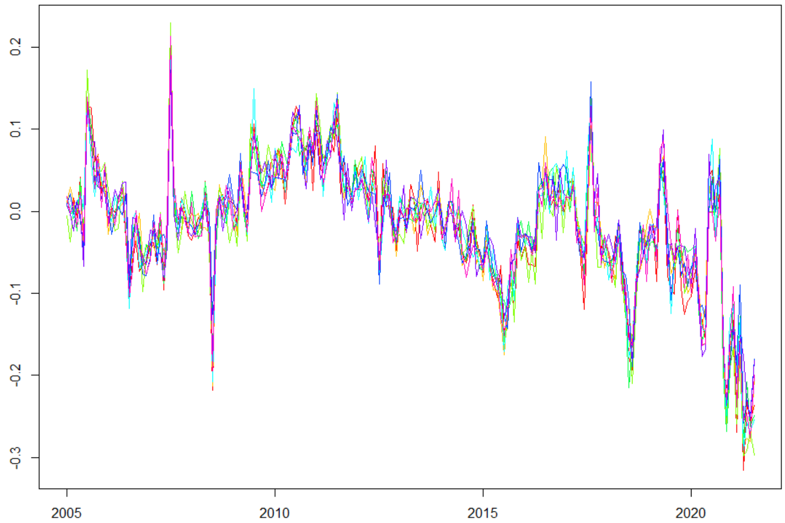


Figure 3.4: Eight Different Samples for “Autos & Vehicles” Google Trends Category

Notes: Google’s methodology introduces measurement error in the data. This plot shows eight different samples retrieved in eight consecutive days for the same Google Trends category (“Autos & Vehicle”), region (U.S.), and time period (2005-2022). It is evident that each sample is slightly different from the other.

regarding the proper treatment of Google Trends data and if they are specified as stationary, trend stationary, or non-stationary. In this chapter we follow Woloszko (2020) and convert them by taking the first seasonal logarithmic difference.

Table 3.1: Correlation Between Different Samples for “Autos & Vehicles”

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Avg. 1	Avg. 2
Day 1	1.00									
Day 2	0.95	1.00								
Day 3	0.95	0.95	1.00							
Day 4	0.94	0.94	0.94	1.00						
Day 5	0.94	0.94	0.94	0.95	1.00					
Day 6	0.94	0.95	0.94	0.94	0.94	1.00				
Day 7	0.92	0.94	0.92	0.92	0.92	0.92	1.00			
Day 8	0.95	0.95	0.95	0.95	0.95	0.94	0.93	1.00		
Avg. 1	0.98	0.98	0.98	0.97	0.96	0.96	0.95	0.97	1.00	
Avg. 2	0.96	0.97	0.96	0.97	0.98	0.97	0.97	0.98	**0.99**	1.00

Our starting point in constructing the Google Trends set of explanatory variables is the selection of primitive keywords. We choose 22 consumption-related keywords that in our view closely track the BEA's National Income and Product Accounts (NIPAs) classification system for Personal Consumption Expenditures (PCE). This framework represents secular trends in consumption patterns owing to the initiation of new products, changes in consumers' preferences, disposable incomes, demographics as well as the greater significance of services. Table 3.3 lists the main components of BEA's classification system for PCE and the primitive keywords that are selected.

Google offers a function called "related queries" in which when a user searches for a specific keyword receives a list of up to 25 related search terms. For example, if a user searches for the keyword "car" on Google Trends, a list of related queries is produced and the top five queries are "car rental", "car wash", "used car", "car insurance", and "rent car". This functionality introduces discipline in our analysis and offers an objective way to construct the set of predictors. Although it is not uncommon in the Google Trends literature to utilise subjective ways to select the predictors (Carrière-Swallow and Labbé, 2013; D'Amuri and Marcucci, 2017; Woloszko, 2020), this is very important since in the first step the selection of primitive keywords conducted on a judgemental basis.

Keyword Planner is free tool offered by Google that assists the user to identify related keywords and other relevant data for digital marketing purposes. We use this tool to identify keywords with the highest average monthly volume and then apply these keywords in Google Trends. We call this dataset as Google Trends "keyword planner". Table B.14 includes all the keywords used to construct this set of predictors. Google Trends "categories" include a broad range of related keywords grouped in a fixed list of categories and follow a 5-level hierarchical classification. They are grouped by using a natural language algorithm whose details are unknown. Table B.15 contains all the categories that are used in this dataset.

Our dataset spans from January 2005 through March 2022 on a monthly frequency. We split the out-of-sample period to examine whether Google Trends data are more useful in periods of low volatility (2017-2019) or in periods of substantial uncertainty (2020-2022).

Table 3.2: Descriptive Statistics

	Mean	Median	S.D.	Min	Max
			Full Sample: 2005-2022		
Durable Goods	0.049	0.055	0.072	-0.195	0.534
Nondurable Goods	0.022	0.023	0.027	-0.077	0.202
Services	0.015	0.019	0.035	-0.210	0.180
PCE	0.020	0.022	0.031	-0.179	0.226
			First Out-of-Sample Period: 2017-2019		
Durable Goods	0.057	0.056	0.021	0.006	0.091
Nondurable Goods	0.026	0.027	0.007	0.002	0.039
Services	0.019	0.018	0.004	0.014	0.029
PCE	0.025	0.024	0.005	0.014	0.034
			Second Out-of-Sample Period: 2020-2022		
Durable Goods	0.104	0.079	0.142	-0.195	0.534
Nondurable Goods	0.056	0.063	0.047	-0.077	0.202
Services	-0.003	0.021	0.094	-0.210	0.180
PCE	0.022	0.027	0.079	-0.179	0.226

Notes: This table shows the following descriptive statistics across the full sample and two sub-periods for personal consumption expenditures (PCE) and its components: mean, media, standard deviation, minimum, and maximum.

3.3 Methodological Framework

This section provides an overview of the methodological framework used in this chapter. Section 3.3.1 introduces the benchmark models that serve as a reference for evaluating the performance of models incorporating Google Trends data. The following sections delve into the specific models employed to evaluate the performance of Google search data. Section 3.3.2 focuses on shrinkage methods, Section 3.3.3 elaborates on factor models, and Section 3.3.4 discusses the ensemble and machine learning models.

3.3.1 Benchmark Models

Autoregressive Model (AR)

To examine whether any information can improve on simply using lags of the target variable itself, we employ as the first benchmark an autoregressive (AR) model of order p:

Table 3.3: PCE Components and Primitive Keywords

BEA's Classification System for PCE	Primitive Keywords
Durable Keywords	
Motor Vehicle & Parts	Car
	Vehicle
Furnishings & Durable Household Equipment	Furnishing
	Home Appliance
Recreational Goods & Vehicles	Sports Car
Other Durables	Computer
	Gaming PC
Nondurable Keywords	
Food & Beverages	Food
	Beverage
Clothing & Footwear	Clothing
Gasoline & Other Energy Goods	Gasoline
	Oil
Other Nondurable Goods	Pharmacy
Services Keywords	
Housing & Utilities	Electricity
Health Care	Health
Transportation Services	Transportation
Recreation Services	Gambling
	Museum
Food Services & Accommodation	Hotel
	Restaurant
Financial Services & Insurance	Insurance
Other Services	Package Delivery

$$C_{i,t} = \mu_i + \sum_{j=1}^{p} \phi_{ij} C_{i,t-j} + u_{i,t} \tag{3.1}$$

where $C_{i,t}$ represents the type of consumption (year-over-year growth rates), μ_i is a constant term, ϕ_{ij} are the autoregressive parameters, $u_{i,t}$ is an error term, and the model order p is chosen so the error term is approximately white noise. The subscript i denotes the different components of private consumption, i.e., durable goods, nondurable goods, and services consumption. For simplicity, in the subsequent sections, we drop the subscript i. The optimal lag order is determined by the Bayesian Information Criterion (BIC), allowing up to four lags, and the parameters are computed by OLS.

Macroeconomic Model (LBC)

To evaluate whether Google Trends variables can provide forecast benefits against that already contained in traditional macroeconomic indicators, we utilise as a second benchmark the model employed by Ludvigson and Bram (1998) and Croushore (2005) (henceforth LBC) in which the following predictors are included: stock prices (measured by the S&P 500 index), interest rates on 3-month Treasury bills (TB3MS) and real personal income (RPI). In addition, we add two measures of sentiment index to the benchmark: the Michigan Consumer Sentiment Index (MCSI) and the Conference Board Consumer Confidence Index (CCI). Thus, the forecasting equation for consumption expenditures and its components has the following form:

$$C_{t+h} = \beta_1 C_{t-1} + \beta_2 SP500_t + \beta_3 TB3MS_t + \beta_4 MCSI_{t-1} + \beta_5 CCI_{t-1} + \beta_6 RPI_{t-2} + \epsilon_{t+h} \quad (3.2)$$

Both sentiment indices attempt to quantify the consumer confidence. The key difference between these indices is that the MCSI index focuses more on financial matters and ongoing plans about significant purchases, while the CCI assigns larger weight on labour market conditions. The different specification of the subscript t indicates that at every prediction only information that was available during that time has been employed. Stock prices and interest rates are available without any lag, while survey and real personal income data are typically published within 30 and 60 days after the end of the reference month, respectively.

Macroeconomic Factor-Augmented Autoregression (M-FAAR)

Give the fact that we live in a data-rich environment, it is reasonable to compare the forecasting performance of Google Trends data against a simple benchmark that uses a large set of macroeconomic predictors. Based on the diffusion index methodology popularised by Stock and Watson (002b), we estimate latent common factors based on 128 economic time series extracted from the FRED-MD macroeconomic database and the prediction at horizon h can be computed as follows:[5]

$$C_{t+h} = \sum_{j=0}^{p} \alpha_j C_{t-j} + \sum_{i=1}^{z} \sum_{j=0}^{k} \beta_{ij} F_{i,t-j} + \epsilon_{t+h} \quad (3.3)$$

[5]For more information regarding the FRED-MD database see McCracken and Ng (2016).

The lag orders of p and k are chosen recursively and sequentially at each prediction horizon using the BIC and allowing up to four lags. The common factors are estimated via principal components applied to a standardised dataset such that $F_t = Ax_t^M$ where A is the rotation matrix and x_t^M contains the macroeconomic series. The number of common factors z has been set to five. The macroeconomic data have been transformed to their stationary form as suggested in McCracken and Ng (2016).

3.3.2 Shrinkage Models

The aim of shrinkage methods is to reduce the dimensionality of the feature space by imposing several penalisation schemes. The coefficients of explanatory variables that do not contain forecasting power for the target variable approach zero or are set equal to zero based on a regularization parameter. Since the empirical literature has not established a clear-cut ranking of the predictive ability of the shrinkage methods, several models are included in our analysis, namely LASSO, adaptive LASSO, and elastic net.

Least Absolute Shrinkage and Selection Operator (LASSO)

The Least Absolute Shrinkage and Selection Operator, introduced by Tibshirani (1996), is a linear regression model, that implements simultaneously regularization and variable selection, and potentially could enhance forecast accuracy and interpretation.

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \left\{ \frac{1}{2} \sum_{t=1}^{T} \left(C_{t+h} - \beta_0 - \sum_{i=1}^{n} \beta_i x_{it}^{GT} \right)^2 + \lambda \sum_{i=1}^{n} |\beta_i| \right\} \tag{3.4}$$

where y_t is the tth observation of the target variable, β_0 is an intercept, x_{ti}^{GT} the tth observation of the ith Google Trends predictor, β_i the corresponding coefficient, and λ is the Lagrangian multiplier or the regularization parameter which controls the amount of shrinkage: when $\lambda = 0$ corresponds to the OLS estimator and when $\lambda \to \infty$ eliminates all coefficients. Essentially, LASSO supports sparse specifications and therefore shrinks the irrelevant predictors to zero. The regularization parameter which controls the amount of shrinkage is determined by the BIC as endorsed by Kock and Callot (2015), Medeiros and Mendes (2016) and Smeekes and Wijler (2018).

Zou (2006) provides evidence that under certain conditions the LASSO's variable selection mechanism could yield inconsistencies. Thus, to achieve model selection consistency the adaptive LASSO (adaLASSO) has been introduced in which adaptive weights are employed for penalizing different coefficients in the $L1$ penalty. The weighted penalty in the adaptive version of LASSO is of the form $\sum_{i=1}^{n} w_i|\beta_i|$ where $w_i = |\hat{\beta}_i|^{\nu}$, $\hat{\beta}_i$ is the OLS estimate, and ν determines how much we want to highlight the differences in the weights. The adaptive LASSO can operate effectively under big data structures and heteroscedasticity.

Elastic Net

While the LASSO estimator is well suited when there is a significant amount of zero coefficients in the true model, there are some limitations that need to be taken into account. Specifically, when there is a strong correlation among predictors, the ridge estimator outperforms the LASSO (Bai and Ng, 008a). Therefore, a generalised approach that incorporates the ridge and the LASSO penalty has been introduced by Zou and Hastie (2005), the elastic net (ElNet):

$$\hat{\beta}^{enet} = \arg\min_{\beta} \left\{ \sum_{t=1}^{T} \left(C_{t+h} - \beta_0 - \sum_{i=1}^{n} \beta_i x_{it}^{GT} \right)^2 + \lambda_1 \sum_{i=1}^{n} |\beta_i| + \lambda_2 \sum_{i=1}^{n} \beta_i^2 \right\} \tag{3.5}$$

Similar to the LASSO estimator, the ElNet performs on the same time shrinkage and variable selection, but also it can choose groups of correlated predictors. The tuning parameters that control the shrinkage, λ_1 and λ_2, are selected using the BIC.

3.3.3 Factor Models

The idea of factor models is to extract latent common factors from a large number of predictors as inputs in forecasting equations which thereby become parsimonious – yet "information rich" – models. This is based on the observation that predictors are often significantly correlated and therefore can be summarised effectively in a small set of factors.

Factor-Augmented Autoregression (FAAR)

We forecast the aggregate consumption and its components by using the diffusion index technique of Stock and Watson (002a), Stock and Watson (002b), and Bai and Ng (2006). This forecasting model estimates the latent common factors, $\hat{F}_t$, and then generates predictions for a target variable, C_{t+h}, based on observed predictors and the common factors. In particular, the forecasts are produced using the following equation for each horizon h:

$$C_{t+h} = \sum_{j=0}^{p} \alpha_j C_{t-j} + \sum_{i=1}^{z} \sum_{j=0}^{k} \beta_{ij} F_{i,t-j} + \epsilon_{t+h} \tag{3.6}$$

The principal components technique is utilised to calculate the common factors such that $F_t = Ax_t^{GT}$, where F_t is a vector of the principal components and A is the rotation matrix. The principal components are applied on a standardised set of stationary series. We follow Stock and Watson (002b) and we determine the number of factors and the lags of the dependent variable and the factors recursively at each forecast horizon using Bayesian Information Criterion (BIC) allowing up to five factors and four lags.

Targeted Factor-Augmented Autoregression (T-FAAR)

As shown by Bai and Ng (008a), including the largest available set of predictors in a forecasting exercise does not necessarily yield superior predictions. The forecasting performance can often be improved by a judicious selection of predictors from which to extract the common factors, including choosing "targeted" predictors, that is, selecting those that have high forecasting power for the target variable.

In this chapter, we adopt the hard-thresholding approach of Bai and Ng (008a) in which predictors are selected according to their association with the target variable. In particular, we regress C_{t+h} against each $x_{j,t}$ predictor where $j = 1, \ldots, N$ as well as autoregressive lags and then extract the t-statistics associated with each $x_{j,t}$. Next, a threshold significance level of $\alpha = 5\%$ is chosen and we select only the statistical significant predictors, $x_t(\alpha)$. Finally, the common factors are estimated from the $x_t(\alpha)$ dataset and the following regression is estimated:

$$C_{t+h} = \sum_{j=0}^{p} \alpha_j C_{t-j} + \sum_{i=1}^{z} \sum_{j=0}^{k} \beta_{ij} f_{i,t-j} + \epsilon_{t+h} \tag{3.7}$$

where $f_t \subseteq F_t$.

3.3.4 Ensemble and Machine Learning Models

Complete Subset Regressions (CSR)

The idea behind Complete Subset Regression (Elliott et al., 2013) is to choose the optimal subset of x_t to predict C_{t+h} by testing all possible combinations of predictors. However, since the computational burden can be substantial, CSR fixes the number of explanatory variables and generates forecast estimations by combining all possible linear regression models. If we have N regressors, CSR chooses $q \leq N$ predictors and estimates all possible combinations q of N. The final prediction is computed as the average across all q-dimensional subsets.

When dealing with high-dimensional datasets CSR might not be the appropriate method. For example, the "related queries" set contains 129 series. If we set $q = 5$ then CSR has to run $254,231,775$ regressions! Hence, we follow a pre-testing procedure such as the one we adopted in the Targeted Factors case in which the target variable C_{t+h} is regressed on each potential candidate $x_{j,t}$, where $j = 1, \ldots, N$, and then we choose the $\tilde{N}$ predictors that possess the highest t-statistic. To minimise the computational workload, we set in our forecasting experiments $\tilde{N} = 20$ and $q = 4$. We experimented with different values for both $\tilde{N}$ and q, but the results showed limited variation.

Random Forests (RF)

Machine learning models are popular methods in statistics and computer science but only recently attracted considerable attention in forecasting economic fundamentals within a data-rich environment. Gogas et al. (2018) use Support Vector Machines to forecast U.S. bank failures; Smeekes and Wijler (2018) utilise LASSO-based procedures to forecast macroeconomic variables with high-dimensional data; Chen et al. (2019) explore the potential ability of machine learning methods and Big Data in forecasting quarterly revenue and expenditure series for the services sector, while Kotchoni et al. (2019) forecast several macroeconomic indicators using numerous high-dimensional and machine learning models. More recently, Medeiros et al. (2021) utilise machine learning models with a large number of covariates to forecast the U.S. inflation, while Babii et al. (2022) introduced structured machine learning regressions for high-dimensional time series data that can be sampled at different frequencies. Findings of the aforementioned studies suggest that machine learning methods can provide significant gains within a macroeconomic forecasting framework.

Machine learning models typically impose mild assumptions about the data-generating process, thus enabling the model to utilise a data-driven approach and to automatically detect the mechanics and intricacies about the underlying statistical relationships in the data. However, the computational burden, the optimal determination of hyper-parameters and the need for sufficient amount of data as well as the meticulous data pre-processing could pose significant issues and limitations for machine learning models.

In this chapter we focus on random forest, a supervised machine learning method introduced by Breiman (2001). It is a non-linear ensemble method based on bootstrap aggregation (bagging) of randomly computed regression trees. The main idea behind regression trees is that they attempt to approximate an unknown non-linear function with local predictions using a binary recursive partitioning procedure. This is an iterative process that separates the predictors into partitions, and then keeps separating each partition into smaller groups until a stopping criterion is reached. Within each region, the predictors are used to model the response by the mean of C_t. Higher order interactions between and within the predictors are automatically included in the random forests since the regression tree might split, for example, according to keywords related to vehicles and furnishings when the target is durable goods expenditures or according to food and energy when we attempt to forecast nondurable goods, and then generate a prediction within each of these partitions.

If we denote by R_m the partition of covariates space where $m = 1, \ldots, M$ and g_m as the node means then the forecast based on regression trees is computed from the following:

$$C_{t+h} = \sum_{m=1}^{M} g_m I_{m(X \in \mathbb{R}, \theta_m)} \tag{3.8}$$

where M is the number of terminal nodes, θ_m is a set of parameters that define the m-th partitioning region, and $I_{m(X \in \mathbb{R}, \theta_m)}$ is an indicator function. Considering that random forests are ensembles of regression trees, the final prediction is equal to the average of the forecasts of each tree:

$$\hat{C}_{t+h} = \frac{1}{B} \sum_{b=1}^{B} \Big[\sum_{m=1}^{M_b} \hat{g}_{m,b} I_{m,b(X \in \mathbb{R}, \hat{\theta}_{m,b})} \Big] \tag{3.9}$$

Each of the regression tree is specified on a bootstrap sub-sample of the original data denoting by b. The number of bootstrap samples is equal to 500. The amount of predictors randomly chosen in each split is 1/3 and remains constant while the forest is grown. Regression trees are grown until there are only five observations in each terminal node.

3.4 Empirical Results

3.4.1 Setup and Main Forecast Exercise

This section presents the main forecasting experiments using Google Trends data and various econometric and machine learning models. Our focus is on forecasting monthly aggregate year-on-year private consumption expenditures (PCE) and its components (durables, nondurables, and services) for the United States across forecasting horizons $h = \{1, \ldots, 9\}$.[6] The out-of-sample forecasting performance of each competing model is

[6] The decision of whether to directly forecast the macroeconomic aggregate or forecast the components first and then aggregate such forecasts remains an empirical question. For example, Marcellino et al.

evaluated relative to a benchmark through the comparison of the root mean squared forecast error (RMSFE). Additionally, we assess the statistical significance of the forecasts by testing for equal predictive accuracy using a Diebold and Mariano (1995) test based on a quadratic loss function. The Diebold and Mariano (DM) test is conducted against a one-sided alternative that Google Trends models produce superior predictions compared to the benchmark.

The sample period covers January 2005 to March 2022. To explore the pandemic's effects and assess the Google Trends data performance under a low- and high-volatility regimes, we divide the out-of-sample period into two sub-periods: 2017:M1-2019:M12 and 2020:M1-2022:M3.[7] In both cases, the estimation sample begins in January 2005. To mitigate potential effects of structural shifts and outliers, we implement a recursive out-of-sample forecasting scheme using a rolling window.[8] As we move forward into the evaluation period, the training sample is extended by one month while the length remains fixed. The models are re-estimated, and a series of forecasts are generated.[9] It is important to note all models use a direct forecasting approach since we do not attempt to forecast the predictors. The only exception is in the case of the autoregressive benchmark in which an iterative approach is employed.

We employ three broad classes of models: shrinkage, factor, and ensemble & machine learning models. The first class of forecasting models comprises elastic net (ElNet), least absolute shrinkage and selection operator (LASSO), and the adaptive version of LASSO. The second class includes the factor-augmented autoregression (FAAR) and the targeted factor (T-FAAR) models. The last category contains complete subset regressions (CSR) and random forests (RF). Each competing model utilises lags of the dependent variable

(2003) provide evidence favouring the aggregation of country-specific models for forecasting inflation in the Euro area, while Hubrich (2005) argues against aggregating forecasts of sub-components of HICP for predicting inflation. For a detailed theoretical discussion on aggregation versus disaggregation in forecasting see, *inter alia*, Giacomini and Granger (2004), Lütkepohl (2006), and Hendry and Hubrich (2011)

[7] The standard deviation for PCE is equal to $\sigma = 0.005$ in the 2017-2019 period, and $\sigma = 0.079$ in the 2020-2022 period.

[8] We also conducted a robustness test using an expanding window, and empirical findings remain very similar.

[9] The length of the rolling window is equal to $T - T_{OOS}$, where T_{OOS} denotes the number of out-of-sample observations.

and Google Trends data as predictors, while benchmark models use autoregressive lags and macroeconomic data. The simple AR model, however, employs only lags of the dependent variable. The number of lags of the target variable is selected using the Bayesian Information Criterion (BIC).

Forecasts are generated for the following variables: durable goods, nondurable goods, services, and the aggregate private consumption expenditure. We initially present the detailed findings using the "related queries" in Tables 3.4 - 3.7 which report the forecast errors. Corresponding empirical findings for Google Trends "keyword planner" and "categories" can be found in Appendix B, specifically in Tables B.1 - B.8.

Panel A of each table exhibits the out-of-sample forecast results for the 2017-2019 period, while Panel B presents findings for the 2020-2022 period. The first three rows of each panel present the actual RMSFEs for the three benchmark models, while the rest show the ratio of the RMSFE of each competing model *vis-à-vis* the RMSFE of the benchmark model that exhibits – in overall – the lowest forecast error, that is, the LBC benchmark model. Thus, values below unity suggest that the corresponding model which utilises Google Trends outperforms the macroeconomic benchmark, while numbers greater than one imply the opposite. For each period under consideration, entries in bold denote the RMSFE-"best" model across all competing Google-based models, for a given forecast horizon. Grey boxes specify the RMSFE-"best" model across the two evaluation periods for a given forecast horizon.

The main findings of the out-of-sample forecasting exercise reveal some interesting insights. First, as depicted in Table 3.4, models incorporating Google Trends seem to underperform the benchmark when forecasting durable expenditures across all prediction horizons during the 2017-2019 period. The underperformance becomes more prominent at $h = 9$. The only exception is the random forest model which outperforms the benchmark at $h = 6$ and $h = 9$. However, a shift to utilising Google Trends "categories" as the main source of predictors results in significant forecast improvements at horizons $h = \{6, 9\}$, particularly when employing random forest and regularization techniques, as illustrated in Table B.5.

Moving to the post-pandemic period, as illustrated in Panel B of Table 3.4, Google-based models appear to outperform the benchmark when forecasting durable goods consumption. Notably, their relative performance improves as the forecast horizon increases, with statistically significant outperformance observed across $h = \{3, 6, 9\}$ as indicated by asterisks in the table. Shrinkage models exhibit the best forecasting performance, showing improvements against the benchmark of up to 16% at $h = 3$, 20% at $h = 6$, and almost 40% at $h = 9$. This pattern is consistently observed when evaluating the performance of "keyword planner" and "categories" Google data in Panel B of Table B.1 and Table B.5, with significant forecast gains predominantly at horizons $h = \{6, 9\}$. Overall, it appears that including a large set of Google Trends predictors combined with shrinkage or random forecast model specifications can lead to significant forecast benefits when predicting durable goods expenditures especially at long-horizons such as six- or nine-months ahead. We hypothesise that Google Trends data might have an edge for predicting durable goods consumption as the nature of these products requires more in-depth research since they are larger investments with a longer lifetime. Also, since it takes a significant amount of time to conduct a research on these products it is also logical that most gains arrive in the medium term (e.g., $h = 9$) and not in the short term (e.g., $h = 1$).

Interesting findings emerge from the forecasting exercise for nondurable goods, as outlined in Table 3.5 when employing "related queries" as predictors. In Panel A, it is evident that before the Covid-19 pandemic, Google-based models offer significant forecast benefits in the short-term and especially at horizon $h = 1$ in which gains reach nearly 17% relative to the benchmark. Additionally, the forecast gains derived from models utilising Google Trends data tend to decline as the prediction horizon increases. Shrinkage and random forecast specifications once again demonstrate strong performance. A similar trend is observed when using "keyword planner" and "categories" predictors, as shown in Table B.2 and Table B.6, where the majority of significant forecast improvements against the benchmark are concentrated at the $h = 1$ horizon. However, when employing Google Trends "categories", the random forest model leads to significant forecast gains across horizons $h = \{3, 6, 9\}$, while combining "keyword planner" data with random forest also yields considerable gains.

In the post-pandemic period, the short-term forecast benefits seem to diminish, as shown in Panel B of Table 3.5, with sporadic improvements primarily at $h = 9$ from shrinkage and random forest models. Interestingly, a similar pattern is observed when utilising predictors derived from "keyword planner", where short-term forecasts exhibit the same trend, but notable gains emerge at horizons $h = \{6, 9\}$ against the benchmark. This pattern is consistent when examining Panel B of Table B.6, revealing gains of up to 15% at $h = 6$ and up to 25% at $h = 9$. Therefore, it seems that significant benefits can also be obtained using Google Trends when forecasting nondurable goods, with these gains largely concentrated at short-term horizons. Once again, shrinkage and random forest models provide the most accurate predictions against the benchmark.

Turning now to services consumption expenditures, the relative out-of-sample performance of Google Trends models remains relatively consistent before and after the pandemic when utilising "related queries", as evidenced in Table 3.6. The 2017-2019 period displays the lowest relative errors across all horizons, denoted by the grey boxes. However, the post-Covid-19 period shows greater statistical improvements, evidenced by the increased amount of times the null hypothesis of the Diebold and Mariano test is rejected. The majority of forecast gains occur at $h = 6$ and $h = 9$. Improvements compared to the benchmark at $h = 6$ reach as high as 19% and 10% before and after the pandemic, respectively, while at $h = 9$, gains are up to 30% and 22%, respectively. Table B.3 reveals that "keyword planner" predictors offer forecast gains at horizons $h = \{6, 9\}$ across both sub-periods, with shrinkage and random forecast models demonstrating the best overall performance. Google "categories" provide negligible forecast benefits.

Finally, Table 3.7 shows that when the main set of predictors comprises "related queries" Google data, significant forecast benefits are evident when forecasting aggregate personal consumption expenditures at horizons $h = \{6, 9\}$ across both the 2017-2019 and 2020-2022 periods. In the pre-pandemic period, the forecast gains are more pronounced, as indicated by the presence of grey boxes in Panel A. Regularization and random forest model specifications demonstrate the best forecasting performance, as suggested by the bold entries. Table B.4 and Table B.8 show a similar pattern when utilising "keyword planner" and "categories" predictors, with the majority of significant forecast improvements against the benchmark located at long-horizons. Shrinkage and random forest models consistently exhibit the lowest relative forecast error compared to the benchmark.

3.4.2 Measuring the Benefits of Google Trends

Having established the predictive power of Google-based models, the next step involves unraveling the source of forecast benefits derived solely from the inclusion of Google Trends data. Therefore, in Tables B.9-B.12, we compare the performance of models utilising Google Trends data against the same models that exclusively incorporate macroeconomic data as predictors. We use the 128 macroeconomic series obtained from the McCracken and Ng (2016) database.[10] Entries below unity imply that Google Trends outperform the macroeconomic data, while values above unity suggest the opposite.

The first panel of Table B.9 shows that "related queries" Google data are a valuable source of information in predicting durable goods expenditures during the 2020-2022 period but not in the period preceding the pandemic. Specifically, it is evident that Google Trends data exhibit a poor forecasting capability when the target variable is durable goods consumption in the 2017-2019 period, as almost all entries are greater than unity, and their performance deteriorates as the forecast horizon increases. However, in the 2020-2022 period, Google search data offer significant forecast gains when included in the set of predictors, as indicated by relative forecast errors which are below unity and the presence of several asterisks, signifying statistical significance according to the Diebold and Mariano test. The second panel of Table B.9 reveals a similar pattern when using the "keyword planner" data, where Google Trends data exhibit poor performance in the pre-pandemic period and strong performance in the 2020-2022 period against macroeconomic data. Additionally, the third panel of Table B.9 shows that "categories" offer some forecast gains not only in the post-pandemic period but also in the 2017-2019

[10] The series have been transformed according to the guidelines of the authors in order to achieve stationarity.

period, mostly at horizons $h = \{6, 9\}$.

Shifting focus to nondurable goods consumption, most entries in the "Related Queries" panel during the 2017-2019 period are close to unity as can be seen in Table B.10, suggesting that the two datasets provide essentially similar predictive power during the period before the pandemic. These findings stand in contrast to those in Table 3.5, which demonstrates a statistically significant outperformance of Google-based models at $h = 1$, indicating that the source of these gains is primarily from the usage of shrinkage, factor, and machine learning models rather than Google Trends indicators. On a similar note, most entries are close to unity during horizons $h = \{1, 2, 3\}$ in the 2020-2022 period, with some numerical benefits emerging at $h = \{6, 9\}$ forecast horizons. Utilising the "keyword planner" and "categories" Google search series yields relative forecast errors close to one during the 2017-2019 period, but some significant gains appear mainly at horizons $h = \{6, 9\}$, as evident in the second and third panel of Table 3.5, respectively.

The Google search "related queries" appear to provide less informative results when directly compared to macroeconomic predictors for forecasting services consumption, as observed in Table B.11. Except for a few cases at the $h = 9$ horizon, the majority of relative forecast errors are close to unity. This pattern holds true when utilising the "keyword planner" and "categories" series as well. Finally, when we forecast the aggregate private consumption, significant gains are obtained using the "related queries" at horizons $h = \{6, 9\}$ in the pre-pandemic period, and a few statistically significant gains emerge at the $h = 9$ horizon in the post-pandemic period, as presented in Table B.12. A similar pattern can be identified when using the "keyword planner" and "categories" series.

3.4.3 Model Performance

Interesting insights can also be revealed when scrutinising the performance of each competing model. Table 3.8 provides a summary regarding the RMSFE–best models. Recall that there are five forecast horizons and two evaluation periods, resulting in 10 specifications for each target variable. Considering four target variables along with three different sets of Google Trends series, a total of 120 forecast estimations are examined. Of the various methods, random forest displays – in general – the lowest forecast errors and secures the top rank in 53 out of 120 forecast estimations. When the set of predictors consists of "related queries" or "keyword planner" predictors the random forest exhibits the best performance in 35% and 37.5% of the cases, respectively, while when "categories" are utilised the random forests shows the best forecasting performance in 60% of the experiments. It is worth noting that the LASSO regression fares quite well across all Google Trends datasets, while targeted factors and complete subset regressions exhibit a relatively disappointing performance.

Valuable insights can be gleaned by delving into the temporal dimension of each model's performance. Figures 3.5-3.8 depict cumulative forecast errors for all Google-based models as well as the best performing benchmark, the LBC model. It is clear that all competing models exhibit almost an equally poor forecasting performance during the first months of the Covid-19 pandemic[11]. Furthermore, it is discernible that two abrupt upward shifts occur in the series. The initial surge in forecast errors corresponds to a stark decline in consumption expenditures during the pandemic's onset, while the second shift is linked to economies reopening swiftly, accompanied by rapid growth and consumption recovery.

3.4.4 Which Google Trends Series Are Selected?

In the concluding section of this chapter, we attempt to identify the Google Trends series most frequently included in sparse settings throughout the entire out-of-sample period. Given that we have already established the substantial forecast gains enabled by Google search data in predicting durable goods at horizons $h = \{3, 6, 9\}$ during the post-pandemic period, as well as in short-horizon predictions for nondurable goods and services, particularly at $h = 9$, it becomes pivotal to gain a more nuanced understanding of the source of predictability in terms of predictor inclusion. Consequently, we examine

[11]This is not surprising. It is well documented by the literature that the accuracy of forecasts from econometric models declines during periods of crisis as well as during steep recoveries. The underperformance during such periods can be attributed to econometric models primarily focusing on capturing the average behaviour of target variables, the dynamic nature of relationships between variables, and the varying impact sizes of economic shocks contingent on crisis nature. For more information see Foroni et al. (2022).

the frequency with which each Google series is selected when employing the LASSO model specification.

Figures 3.9 - 3.11 show the frequency of the top 10 Google Trends inclusions over time for horizons $h = \{1, 6, 9\}$. A red box denotes that the corresponding predictor has been included in the forecasting regression during a specific out-of-sample prediction. Figure 3.9 shows the most frequently selected Google Trends series throughout the out-of-sample period when the target variable is durable goods consumption expenditures. At horizon $h = 1$, the most popular series is "sports car", included in the LASSO forecast regressions in over 60% of cases. For horizons $h = 6$ and $h = 9$, "home depot appliance" and "lowes" emerge as the most frequent Google Trends predictors, respectively. Overall, it seems that a particular set of keywords ("lowes", "home furnishing stores", "vehicle registration") is loaded in the pre-pandemic period, while a distinct set of keywords, primarily related to computers and gaming, takes precedence in the post-pandemic period.

Next, Figure 3.10 shows the inclusion of Google search data when forecasting non-durable goods consumption and it seems that the most frequently selected query across all horizons is "walgreens", selected around in 65-70% of the monthly predictions. Keywords related to gasoline, beverages, and clothing have also a significant presence in the model across all forecast horizons. Moreover, it is noteworthy that Google Trends series are included more often compared to the other target variables, with a notable surge in selection observed from 2017 to 2020.

Finally, Figure 3.11 reveals the frequency with which the Google data are loaded in the LASSO specification when the target variable is services consumption. At horizon $h = 1$, the most frequent predictor is "public health", entering the model more than 60% of iterations, mostly during the 2019-2022 period. As we progress to forecast horizons $h = \{6, 9\}$, "transportation" becomes more prominent, employed in the model 38 out of 63 monthly iterations during the out-of-sample period. Several keywords related to electricity and gambling are also included a considerable number of times. Additionally, it is observed that as the forecast horizon increases from $h = 1$ to $h = 9$, the predictors are loaded into the model more frequently.

Overall, during the out-of-sample period, the process of variable selection exhibits dynamic characteristics influenced by the target variable and forecast horizon. Importantly, the analysis reveals a temporal shift in variable inclusion patterns. Specifically, pre-pandemic, a greater number of variables found their way into the model compared to the period following 2020. This phenomenon suggests that the predictive ability of Google Trends series can change over time depending on the business cycle.

3.5 Concluding Remarks

This empirical chapter demonstrates that integrating a large dataset of Google search data into the forecaster's toolkit, along with the application of shrinkage and random forest models, can result in substantial improvements in forecasting accuracy for consumption. Specifically, the utilisation of Google Trends "related queries" leads to forecast benefits for aggregate U.S. private consumption expenditures at horizons $h = \{3, 6, 9\}$, with performance gains ranging approximately from 15% to 20% compared to the benchmark model.

Additionally, we observe considerable improvements when forecasting distinct components of Personal Consumption Expenditure (PCE). In particular, durable goods expenditures benefit notably from the incorporation of Google-based models in the post-pandemic era, particularly at horizons $h = \{3, 6, 9\}$, with gains of approximately 15-40% relative to the benchmark. We also document significant forecast benefits in the 2017-2019 period when forecasting nondurable goods expenditures mostly at horizons $h = \{1, 2\}$. Service expenditures exhibit noticeable gains at the 9-month horizon, with forecast gains of up to 30% over the benchmark.

Furthermore, a comparison of the performance of each competing model using only macroeconomic data against the same model incorporating Google Trends data as predictors yields some interesting insights. For durable goods expenditures, Google Trends data significantly enhance forecasts, especially at $h = \{3, 6, 9\}$, with gains of up to 50% compared to macroeconomic predictors. Conversely, nondurable goods demonstrate comparable performance between Google Trends and macroeconomic-based models, except for a few cases at horizon $h = 9$. This implies that the substantial gains documented compared to the benchmark are primarily attributable to the utilisation of advanced forecasting models rather than Google Trends data. Sporadic gains appear for services mostly at $h = 9$ horizons, while for aggregate consumption expenditures significant gains are documented primarily at horizons $h = \{6, 9\}$.

We find that the models that impose sparsity provide the highest forecast accuracy. By contrast, specifications that utilise a dense approach such as factor models perform quite poorly during the out-of-sample exercise. Overall, the model with the highest forecast accuracy across the different Google Trends datasets, target variables, and horizons has been the random forest which has reached the top more than 40% of our forecast experiments. LASSO regression also performs well, while complete subset regressions show weaker performance. In terms of the temporal dimension, all models face severe challenges in accuracy during the early stages of the Covid-19 pandemic, with notable spikes in forecast errors during the initial decline in consumption expenditures and subsequent rapid economic recovery.

In the last part of our empirical analysis, we investigate the selection of Google Trends series during the complete out-of-sample period, focusing on their frequency in sparse settings for various forecasting horizons and target variables. Our analysis uncovers dynamic trends in variable inclusion patterns, revealing a temporal shift in predictive factors. Pre-pandemic, a greater number of variables found their way into the model compared to the period following 2020, signifying a changing landscape of predictive abilities over time, potentially tied to the business cycle.

Table 3.4: Forecasting Durable Goods Consumption using "Related Queries"

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
	Panel A: 2017-2019				
AR	0.015	0.016	0.018	0.024	0.027
LBC	0.015	0.017	0.018	0.023	0.026
M-FAAR	0.013	0.017	0.019	0.024	0.028
ElNet	0.989	1.093	1.174	1.178	1.320
LASSO	**0.976**	1.051	1.227	1.185	1.342
adaLASSO	0.985	1.050	1.148	1.281	1.479
FAAR	1.016	**1.013**	1.094	1.171	1.141
T-FAAR	1.016	1.015	1.131	1.220	1.250
CSR	1.061	1.065	**1.094**	1.193	1.387
RF	1.096	1.130	1.108	**0.854****	**0.777***
	Panel B: 2020-2022				
AR	0.117	0.157	0.168	0.171	0.182
LBC	0.111	0.152	0.168	0.169	0.175
M-FAAR	0.150	0.182	0.194	0.194	0.186
ElNet	0.984	0.891	**0.841***	0.815**	0.661**
LASSO	**0.946**	**0.878***	0.849*	0.819**	0.659**
adaLASSO	0.950	0.909*	0.873**	0.855**	**0.627****
FAAR	1.031	0.931	0.890**	0.856**	0.785*
T-FAAR	1.054	0.926*	0.894**	0.860**	0.846*
CSR	1.035	0.982	0.932	0.911	0.812*
RF	1.180	0.946	0.849**	**0.801****	0.809*

Notes: The first three rows in each panel show the root mean squared forecast error (RMSFE) for the benchmark models (AR, LBC, M-FAAR). The rest of the entries show the RMSFE of each model relative to the best performing benchmark which is the LBC model. Thus, values lower than unity suggest a particular model which utilises only Google Trends data outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon. Grey boxes denote the lowest relative error achieved by the competing models between the two subperiods. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table 3.5: Forecasting Nondurable Goods Consumption using "Related Queries"

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
	Panel A: 2017-2019				
AR	0.008	0.008	0.009	0.010	0.011
LBC	0.009	0.009	0.009	0.010	0.010
M-FAAR	0.008	0.009	0.009	0.010	0.010
ElNet	0.840*	0.900	0.912	0.971	0.933
LASSO	**0.836***	**0.885***	0.914	0.978	0.968
adaLASSO	0.882*	0.893*	0.911*	0.998	0.932
FAAR	0.856*	0.935	**0.900***	0.940	1.041
T-FAAR	0.890	1.059	1.025	1.131	1.111
CSR	0.845*	0.919	0.917	1.013	1.108
RF	0.874*	0.930	0.907*	**0.936**	**0.917**
	Panel B: 2020-2022				
AR	0.050	0.054	0.056	0.058	0.073
LBC	0.048	0.051	0.053	0.057	0.062
M-FAAR	0.067	0.058	0.064	0.061	0.066
ElNet	**0.987**	**0.983**	**0.944**	0.905	**0.810****
LASSO	0.996	1.015	1.012	0.922	0.818**
adaLASSO	1.006	1.062	1.108	0.931	0.838*
FAAR	1.001	1.124	1.038	1.005	0.962
T-FAAR	1.013	1.017	1.112	1.045	1.022
CSR	1.054	1.042	1.025	0.950	1.047
RF	1.052	0.984	0.971	**0.898***	0.869*

Notes: The first three rows in each panel show the root mean squared forecast error (RMSFE) for the benchmark models (AR, LBC, M-FAAR). The rest of the entries show the RMSFE of each model relative to the best performing benchmark which is the LBC model. Thus, values lower than unity suggest a particular model which utilises only Google Trends data outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon. Grey boxes denote the lowest relative error achieved by the competing models between the two subperiods. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table 3.6: Forecasting Services Goods Consumption using "Related Queries"

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
	Panel A: 2017-2019				
AR	0.003	0.003	0.004	0.005	0.006
LBC	0.003	0.003	0.004	0.005	0.005
M-FAAR	0.003	0.003	0.004	0.004	0.004
ElNet	**0.935**	1.007	0.995	0.937	0.738*
LASSO	0.985	1.004	0.991	0.922	0.897
adaLASSO	1.003	0.990	1.001	0.920	0.753*
FAAR	1.032	1.016	1.040	0.982	0.845
T-FAAR	1.013	0.987	0.999	0.936	1.179
CSR	0.982	**0.966**	**0.970**	1.016	0.979
RF	1.092	1.122	1.081	**0.812***	**0.694****
	Panel B: 2020-2022				
AR	0.082	0.096	0.093	0.097	0.089
LBC	0.059	0.080	0.087	0.094	0.097
M-FAAR	0.066	0.087	0.095	0.108	0.105
ElNet	1.103	1.034	1.019	0.995	0.837*
LASSO	**1.060**	**1.025**	1.001	0.997	0.802*
adaLASSO	1.066	1.062	1.017	0.985	**0.777***
FAAR	1.292	1.166	1.026	0.997	1.007
T-FAAR	1.203	1.115	**0.990**	1.011	0.844*
CSR	1.382	1.062	1.008	0.925*	0.782**
RF	1.142	1.041	1.046	**0.909***	0.778**

Notes: The first three rows in each panel show the root mean squared forecast error (RMSFE) for the benchmark models (AR, LBC, M-FAAR). The rest of the entries show the RMSFE of each model relative to the best performing benchmark which is the LBC model. Thus, values lower than unity suggest a particular model which utilises only Google Trends data outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon. Grey boxes denote the lowest relative error achieved by the competing models between the two subperiods. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table 3.7: Forecasting Private Consumption using "Related Queries"

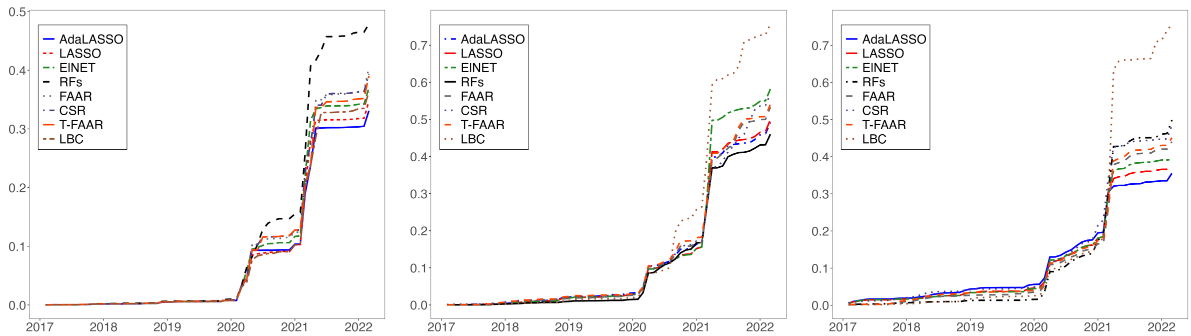
	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
	Panel A: 2017-2019				
AR	0.004	0.005	0.005	0.007	0.008
LBC	0.004	0.005	0.005	0.006	0.006
M-FAAR	0.004	0.005	0.005	0.006	0.006
ElNet	**0.942**	**0.921***	0.935	0.831*	0.843
LASSO	1.014	0.945	0.982	0.836*	**0.773***
adaLASSO	1.007	0.984	0.949	0.848*	0.777*
FAAR	0.993	0.967	1.034	1.141	1.138
T-FAAR	1.028	1.065	1.121	1.195	1.220
CSR	1.009	0.958	0.947	0.947	0.886*
RF	0.994	0.949	**0.919***	**0.791****	0.864*
	Panel B: 2020-2022				
AR	0.076	0.083	0.081	0.084	0.086
LBC	0.055	0.073	0.078	0.081	0.080
M-FAAR	0.063	0.079	0.087	0.094	0.089
ElNet	1.193	1.002	**0.976**	0.948*	0.858*
LASSO	1.192	0.998	0.988	0.919*	0.831*
adaLASSO	**1.127**	1.037	1.006	0.908**	0.815*
FAAR	1.315	1.076	1.011	0.961	0.891*
T-FAAR	1.332	1.063	1.014	1.065	0.983
CSR	1.392	1.145	1.003	0.917	0.890*
RF	1.185	**0.994**	0.989	**0.891****	**0.808****

Notes: The first three rows in each panel show the root mean squared forecast error (RMSFE) for the benchmark models (AR, LBC, M-FAAR). The rest of the entries show the RMSFE of each model relative to the best performing benchmark which is the LBC model. Thus, values lower than unity suggest a particular model which utilises only Google Trends data outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon. Grey boxes denote the lowest relative error achieved by the competing models between the two subperiods. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table 3.8: Summary of the RMSFE-best Models Across All Google Trends Datasets

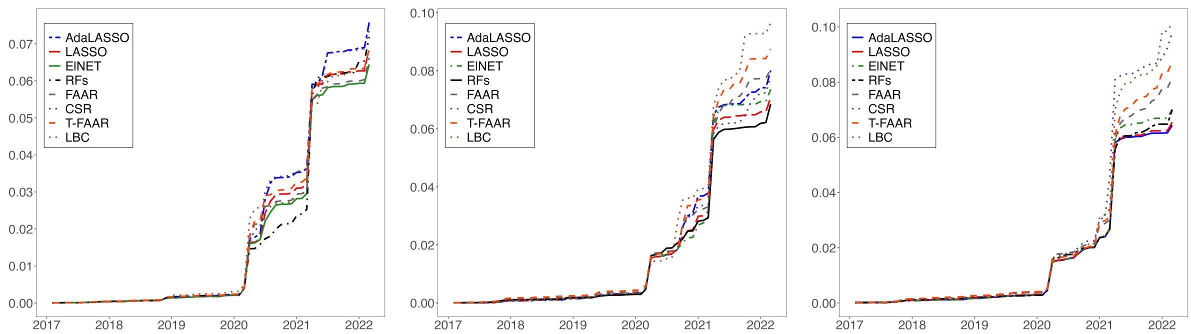
	Related Queries	Keyword Planner	Categories	Total
ElNet	9	8	3	20
LASSO	8	10	7	25
adaLASSO	3	1	4	8
FAAR	2	5	0	7
T-FAAR	1	0	2	3
CSR	3	1	0	4
RF	14	15	24	53
Total	40	40	40	120

Notes: Entries in this table show the number of times that each forecasting model produces the lowest relative error compared to the benchmark, for a given set of Google search data. There are five estimations for each horizon, two evaluating periods, and four target variables, meaning 40 forecast estimations for each Google Trends dataset. Thus, in total there are 120 estimations.



(a) Horizon = 1 (b) Horizon = 6 (c) Horizon = 9

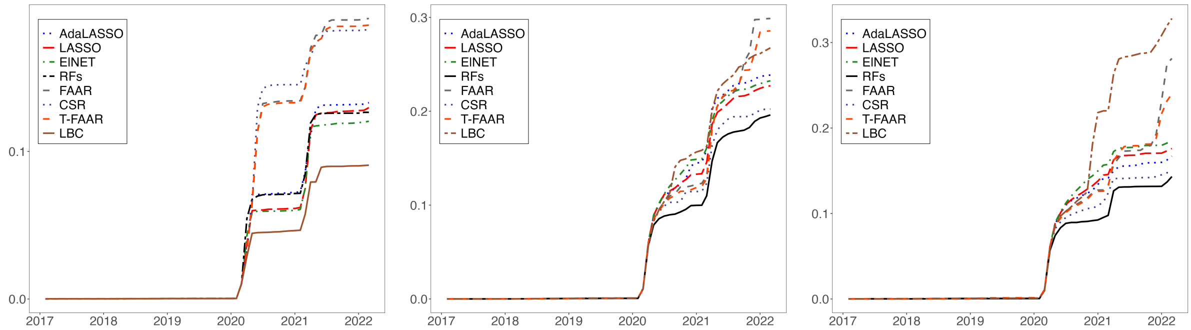
Figure 3.5: Cumulative RMSFE: Durables



(a) Horizon = 1 (b) Horizon = 6 (c) Horizon = 9

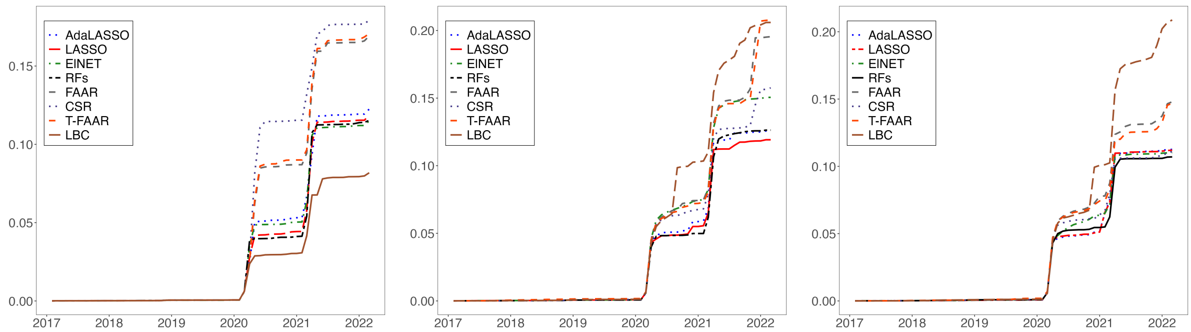
Figure 3.6: Cumulative RMSFE: Nondurables

Notes: This graph shows the cumulative root mean squared forecast errors (RMSFE). Panel (a) exhibits rolling RMSFE for one-step-ahead forecasts, panel (b) for six-step-ahead forecasts, and panel (c) for nine-step-ahead forecasts. In each graph, the model that displays the lowest cumulative error is depicted in a solid line while the rest appear in dashed and dotted lines.



(a) Horizon = 1 (b) Horizon = 6 (c) Horizon = 9

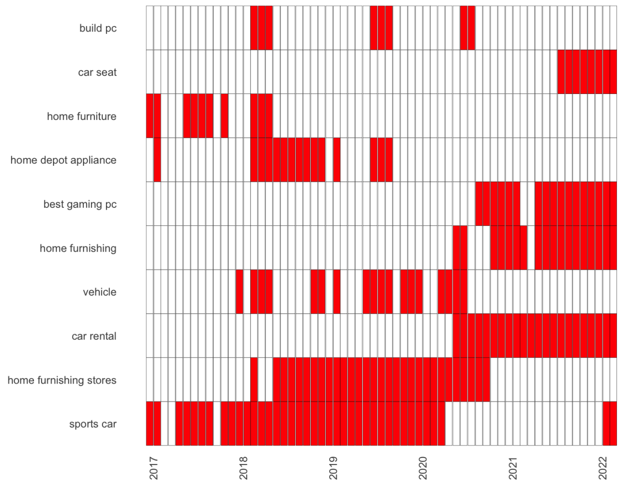
Figure 3.7: Cumulative RMSFE: Services



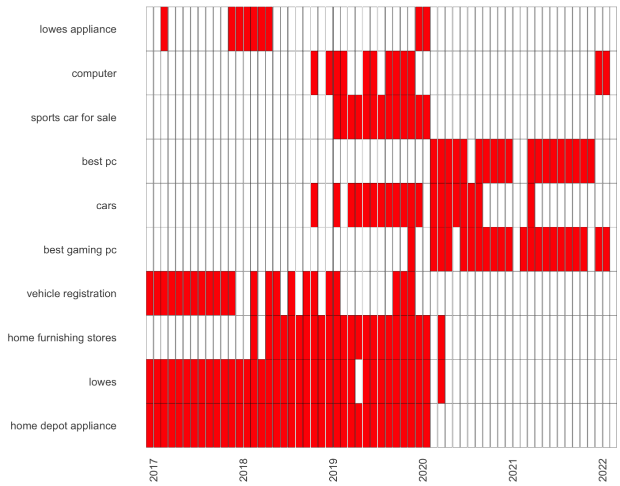
(a) Horizon = 1 (b) Horizon = 6 (c) Horizon = 9

Figure 3.8: Cumulative RMSFE: PCE

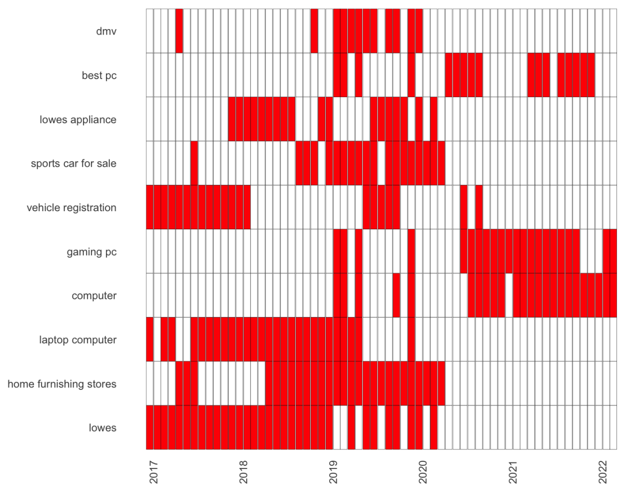
Notes: This graph shows the cumulative root mean squared forecast errors (RMSFE). Panel (a) exhibits rolling RMSFE for one-step-ahead forecasts, panel (b) for six-step-ahead forecasts, and panel (c) for nine-step-ahead forecasts. In each graph, the model that displays the lowest cumulative error is depicted in a solid line while the rest appear in dashed and dotted lines.



(a) Horizon = 1



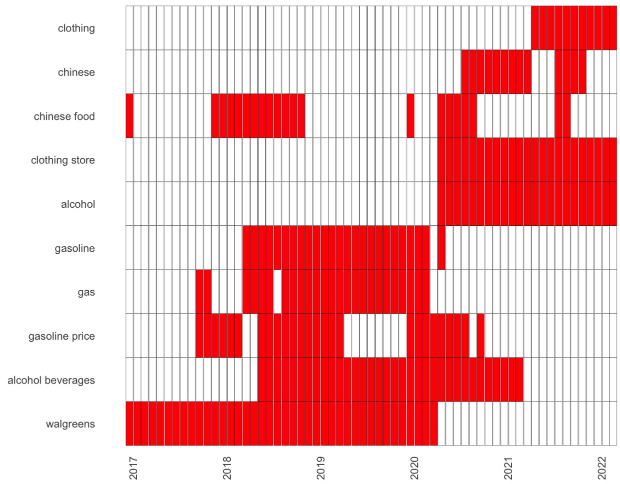
(b) Horizon = 6



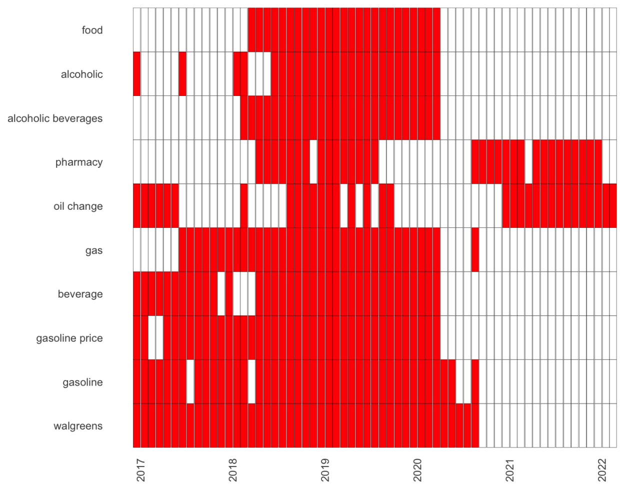
(c) Horizon = 9

Figure 3.9: Predictor Inclusion in LASSO Regressions: Durable Goods

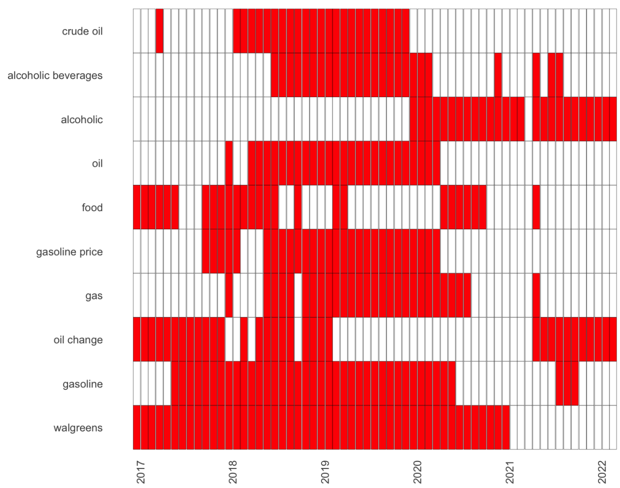
Notes: This graph shows the selection of the top 10 Google Trends predictors in LASSO regressions over the full out-of-sample period when forecasting durable goods consumption. The variables are ordered from bottom to top according to their inclusion frequency during the out-of-sample period.



(a) Horizon = 1



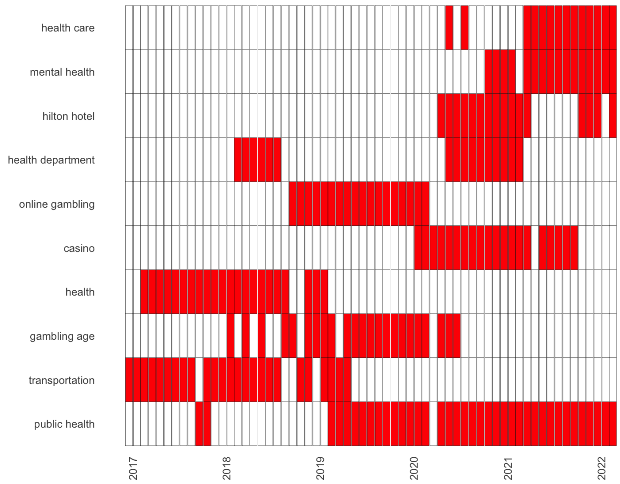
(b) Horizon = 6



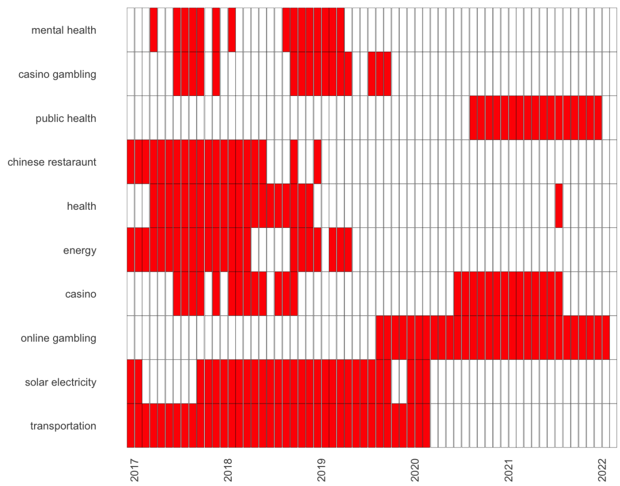
(c) Horizon = 9

Figure 3.10: Predictor Inclusion in LASSO Regressions: Nondurable Goods

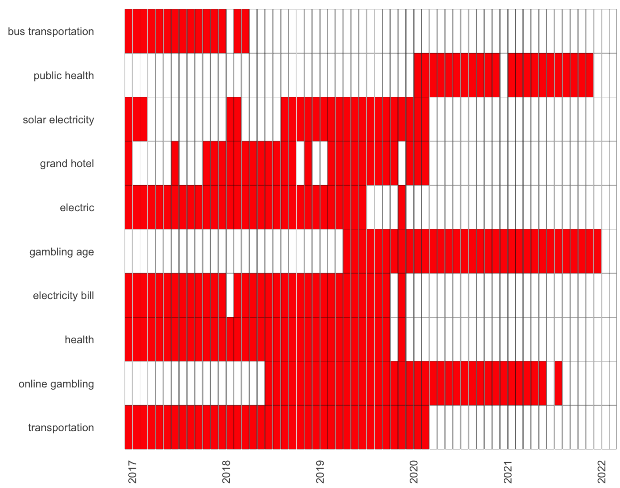
Notes: This graph shows the selection of the top 10 Google Trends predictors in LASSO regressions over the full out-of-sample period when forecasting nondurable goods consumption. The variables are ordered from bottom to top according to their inclusion frequency during the out-of-sample period.



(a) Horizon = 1



(b) Horizon = 6



(c) Horizon = 9

Figure 3.11: Predictor Inclusion in LASSO Regressions: Services

Notes: This graph shows the selection of the top 10 Google Trends predictors in LASSO regressions over the full out-of-sample period when forecasting services consumption. The variables are ordered from bottom to top according to their inclusion frequency during the out-of-sample period.

B Appendix

Table B.1: Forecasting Durable Goods Consumption using "Keyword Planner"

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
			Panel A: 2017-2019		
AR	0.015	0.016	0.018	0.024	0.027
LBC	0.015	0.017	0.018	0.023	0.026
M-FAAR	0.013	0.017	0.019	0.024	0.028
ElNet	1.002	1.138	1.186	1.190	1.021
LASSO	**0.955**	1.095	1.145	1.225	1.070
adaLASSO	0.978	1.034	1.141	1.277	1.197
FAAR	0.997	**0.967**	**0.901***	**0.937**	**0.778***
T-FAAR	0.968	1.001	1.055	1.176	1.325
CSR	1.015	1.157	1.254	1.487	1.534
RF	1.021	1.048	1.030	1.197	1.425
			Panel B: 2020-2022		
AR	0.117	0.157	0.168	0.171	0.182
LBC	0.111	0.152	0.168	0.169	0.175
M-FAAR	0.150	0.182	0.194	0.194	0.186
ElNet	1.009	**0.862***	0.930	0.811**	0.776*
LASSO	**0.975**	0.883	0.909	**0.783****	**0.762***
adaLASSO	0.991	0.961	0.985	0.929	0.767*
FAAR	1.018	0.925*	0.891**	0.844**	0.847
T-FAAR	1.040	0.975	0.979	0.984	0.890*
CSR	1.040	1.003	0.955	0.920	0.885
RF	1.137	0.903*	**0.837****	0.800**	0.796*

Notes: The first three rows in each panel show the root mean squared forecast error (RMSFE) for the benchmark models (AR, LBC, M-FAAR). The rest of the entries show the RMSFE of each model relative to the best performing benchmark which is the LBC model. Thus, values lower than unity suggest a particular model which utilises only Google Trends data outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon. Grey boxes denote the lowest relative error achieved by the competing models between the two subperiods. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table B.2: Forecasting Nondurable Goods Consumption using "Keyword Planner"

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
			Panel A: 2017-2019		
AR	0.008	0.008	0.009	0.010	0.011
LBC	0.009	0.009	0.009	0.010	0.010
M-FAAR	0.008	0.009	0.009	0.010	0.010
ElNet	0.860*	0.952	0.927	0.964	0.983
LASSO	**0.836***	0.935	0.932	0.909	1.026
adaLASSO	0.880*	0.936	0.928	0.942	1.020
FAAR	0.907	0.966	0.920	0.977	**0.810****
T-FAAR	0.841*	0.938	0.902*	0.954	0.913
CSR	0.839*	**0.904***	0.884*	0.912	0.824*
RF	0.842*	0.920	**0.863**	**0.849**	0.828*
			Panel B: 2020-2022		
AR	0.050	0.054	0.056	0.058	0.073
LBC	0.048	0.051	0.053	0.057	0.062
M-FAAR	0.067	0.058	0.064	0.061	0.066
ElNet	0.998	**0.937**	**0.900**	**0.840****	0.793**
LASSO	1.000	0.946	0.963	0.852**	**0.779****
adaLASSO	1.049	1.011	1.076	0.904	0.789*
FAAR	0.990	0.992	0.975	0.926	0.832
T-FAAR	0.979	0.966	1.012	0.921*	0.875*
CSR	1.029	0.994	0.977	0.880*	0.899
RF	**0.971**	0.951	0.942	0.851**	0.783**

Notes: The first three rows in each panel show the root mean squared forecast error (RMSFE) for the benchmark models (AR, LBC, M-FAAR). The rest of the entries show the RMSFE of each model relative to the best performing benchmark which is the LBC model. Thus, values lower than unity suggest a particular model which utilises only Google Trends data outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon. Grey boxes denote the lowest relative error achieved by the competing models between the two subperiods. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table B.3: Forecasting Services Goods Consumption using "Keyword Planner"

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
	Panel A: 2017-2019				
AR	0.003	0.003	0.004	0.005	0.006
LBC	0.003	0.003	0.004	0.005	0.005
M-FAAR	0.003	0.003	0.004	0.004	0.004
ElNet	**0.904***	**0.912***	**0.926**	0.847*	0.830
LASSO	0.974	0.952	0.949	0.846**	0.800*
adaLASSO	1.003	0.978	1.003	0.879**	0.794*
FAAR	1.035	1.019	1.019	0.934	0.918
T-FAAR	1.015	0.996	1.014	1.077	1.329
CSR	1.003	0.927	0.957	0.982	0.941
RF	1.090	1.122	1.086	**0.817****	**0.699****
	Panel B: 2020-2022				
AR	0.082	0.096	0.093	0.097	0.089
LBC	0.059	0.080	0.087	0.094	0.097
M-FAAR	0.066	0.087	0.095	0.108	0.105
ElNet	**1.016**	1.000	1.002	0.982	0.855*
LASSO	1.036	**0.974**	**1.000**	0.980	0.833*
adaLASSO	1.068	0.999	1.031	0.973	0.830*
FAAR	1.335	1.102	1.029	0.999	0.950
T-FAAR	1.306	1.094	1.001	0.991	0.867*
CSR	1.356	1.159	1.020	0.944*	0.797**
RF	1.170	1.036	1.009	**0.919***	**0.791****

Notes: The first three rows in each panel show the root mean squared forecast error (RMSFE) for the benchmark models (AR, LBC, M-FAAR). The rest of the entries show the RMSFE of each model relative to the best performing benchmark which is the LBC model. Thus, values lower than unity suggest a particular model which utilises only Google Trends data outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon. Grey boxes denote the lowest relative error achieved by the competing models between the two subperiods. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table B.4: Forecasting Private Consumption using "Keyword Planner"

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
			Panel A: 2017-2019		
AR	0.004	0.005	0.005	0.007	0.008
LBC	0.004	0.005	0.005	0.006	0.006
M-FAAR	0.004	0.005	0.005	0.006	0.006
ElNet	0.973	0.950	0.951	0.860*	0.959
LASSO	1.030	0.944	0.940	0.834*	0.979
adaLASSO	1.008	0.983	0.906	0.948	0.988
FAAR	1.034	1.041	1.005	0.967	0.831*
T-FAAR	1.027	1.028	1.054	1.119	1.136
CSR	1.004	0.965	0.959	1.009	1.063
RF	**0.967**	**0.877***	**0.836***	**0.761***	**0.743****
			Panel B: 2020-2022		
AR	0.076	0.083	0.081	0.084	0.086
LBC	0.055	0.073	0.078	0.081	0.080
M-FAAR	0.063	0.079	0.087	0.094	0.089
ElNet	1.083	0.987	0.955*	0.959	0.837*
LASSO	**1.035**	0.990	**0.898***	0.940*	0.809*
adaLASSO	1.087	1.021	0.933*	0.945*	**0.783****
FAAR	1.308	1.057	1.008	0.958*	0.854*
T-FAAR	1.309	1.056	0.981	0.940*	0.820**
CSR	1.392	1.099	0.978	0.910*	0.908
RF	1.169	**0.966***	0.902*	**0.834****	0.801**

Notes: The first three rows in each panel show the root mean squared forecast error (RMSFE) for the benchmark models (AR, LBC, M-FAAR). The rest of the entries show the RMSFE of each model relative to the best performing benchmark which is the LBC model. Thus, values lower than unity suggest a particular model which utilises only Google Trends data outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon. Grey boxes denote the lowest relative error achieved by the competing models between the two subperiods. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table B.5: Forecasting Durable Goods Consumption using “Categories”

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
	Panel A: 2017-2019				
AR	0.015	0.016	0.018	0.024	0.027
LBC	0.015	0.017	0.018	0.023	0.026
M-FAAR	0.013	0.017	0.019	0.024	0.028
ElNet	0.970	0.982	1.014	0.871	0.746**
LASSO	**0.969**	**0.942**	0.999	0.911	0.759**
adaLASSO	0.993	0.996	1.089	1.028	0.846*
FAAR	0.994	0.983	0.949	0.988	0.955
T-FAAR	1.019	1.062	1.064	1.171	1.157
CSR	1.018	0.991	1.003	0.974	0.816
RF	1.027	0.970	**0.911**	**0.608***	**0.631****
	Panel B: 2020-2022				
AR	0.117	0.157	0.168	0.171	0.182
LBC	0.111	0.152	0.168	0.169	0.175
M-FAAR	0.150	0.182	0.194	0.194	0.186
ElNet	1.111	0.957	0.942	0.809**	0.734**
LASSO	1.033	**0.937**	0.929	0.792**	0.735**
adaLASSO	**1.026**	1.004	1.015	0.810**	**0.707****
FAAR	1.051	0.983	0.985	0.931*	0.851*
T-FAAR	1.028	0.981	0.948	0.914*	0.907
CSR	1.075	1.036	1.001	0.953	0.947
RF	1.247	0.983	**0.921**	**0.788****	0.780**

Notes: The first three rows in each panel show the root mean squared forecast error (RMSFE) for the benchmark models (AR, LBC, M-FAAR). The rest of the entries show the RMSFE of each model relative to the best performing benchmark which is the LBC model. Thus, values lower than unity suggest a particular model which utilises only Google Trends data outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon. Grey boxes denote the lowest relative error achieved by the competing models between the two subperiods. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table B.6: Forecasting Nondurable Goods Consumption using "Categories"

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
	Panel A: 2017-2019				
AR	0.008	0.008	0.009	0.010	0.011
LBC	0.009	0.009	0.009	0.010	0.010
M-FAAR	0.008	0.009	0.009	0.010	0.010
ElNet	0.879*	0.941	0.947	0.994	0.941
LASSO	0.886*	0.944	0.928	1.011	0.978
adaLASSO	0.877*	0.972	0.936	1.003	0.971
FAAR	0.923	1.038	1.010	1.113	1.197
T-FAAR	0.887	1.040	0.975	1.072	1.069
CSR	0.872*	0.936	0.924	0.921	0.903
RF	**0.855***	**0.911**	**0.877***	**0.840***	**0.779****
	Panel B: 2020-2022				
AR	0.050	0.054	0.056	0.058	0.073
LBC	0.048	0.051	0.053	0.057	0.062
M-FAAR	0.067	0.058	0.064	0.061	0.066
ElNet	**0.888***	0.953	0.895	0.845**	**0.762***
LASSO	0.931	0.971	0.945	0.852**	0.775*
adaLASSO	0.956	1.018	0.996	0.859**	0.799*
FAAR	0.931*	0.971	0.945	0.852**	0.775*
T-FAAR	1.072	1.035	1.071	0.962	0.931
CSR	1.035	1.002	0.963	0.926	0.990
RF	0.973	**0.922**	**0.889***	**0.844****	0.837*

Notes: The first three rows in each panel show the root mean squared forecast error (RMSFE) for the benchmark models (AR, LBC, M-FAAR). The rest of the entries show the RMSFE of each model relative to the best performing benchmark which is the LBC model. Thus, values lower than unity suggest a particular model which utilises only Google Trends data outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon. Grey boxes denote the lowest relative error achieved by the competing models between the two subperiods. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table B.7: Forecasting Services Goods Consumption using “Categories”

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
		Panel A: 2017-2019			
AR	0.003	0.003	0.004	0.005	0.006
LBC	0.003	0.003	0.004	0.005	0.005
M-FAAR	0.003	0.003	0.004	0.004	0.004
ElNet	**0.958**	1.102	1.248	1.229	1.314
LASSO	1.002	1.008	1.120	1.184	1.314
adaLASSO	1.004	1.031	1.104	1.204	1.335
FAAR	1.049	1.043	1.068	1.175	1.589
T-FAAR	1.017	**1.002**	**1.037**	1.088	1.176
CSR	1.024	1.026	1.045	1.008	0.967
RF	1.065	1.102	1.054	**0.779***	**0.704***
		Panel B: 2020-2022			
AR	0.082	0.096	0.093	0.097	0.089
LBC	0.059	0.080	0.087	0.094	0.097
M-FAAR	0.066	0.087	0.095	0.108	0.105
ElNet	1.058	1.021	1.007	0.941	0.838*
LASSO	1.035	1.013	0.974	0.902	0.787**
adaLASSO	**1.023**	1.066	1.007	0.906	0.777**
FAAR	1.480	1.039	1.014	0.980	1.068
T-FAAR	1.323	1.083	0.979	0.957	0.942
CSR	1.324	1.133	0.991	0.891*	0.883*
RF	1.184	**0.947**	**0.869***	**0.771****	**0.720****

Notes: The first three rows in each panel show the root mean squared forecast error (RMSFE) for the benchmark models (AR, LBC, M-FAAR). The rest of the entries show the RMSFE of each model relative to the best performing benchmark which is the LBC model. Thus, values lower than unity suggest a particular model which utilises only Google Trends data outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon. Grey boxes denote the lowest relative error achieved by the competing models between the two subperiods. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table B.8: Forecasting Private Consumption using "Categories"

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
			Panel A: 2017-2019		
AR	0.004	0.005	0.005	0.007	0.008
LBC	0.004	0.005	0.005	0.006	0.006
M-FAAR	0.004	0.005	0.005	0.006	0.006
ElNet	1.009	1.015	0.936	0.969	1.179
LASSO	1.012	0.969	**0.904***	1.075	**0.868***
adaLASSO	1.008	0.999	0.969	1.097	1.004
FAAR	1.007	0.999	1.000	1.104	1.159
T-FAAR	1.011	1.021	1.014	1.135	1.218
CSR	1.033	1.008	1.043	1.156	1.194
RF	**0.979**	**0.942**	0.909	**0.923**	1.026
			Panel B: 2020-2022		
AR	0.076	0.083	0.081	0.084	0.086
LBC	0.055	0.073	0.078	0.081	0.080
M-FAAR	0.063	0.079	0.087	0.094	0.089
ElNet	1.103	0.982	0.978	0.886*	0.874*
LASSO	**1.067**	0.975	0.954*	**0.808****	0.820*
adaLASSO	1.108	1.017	0.982	0.826**	**0.819****
FAAR	1.331	1.083	1.046	1.012	1.006
T-FAAR	1.324	1.018	1.015	0.987	1.032
CSR	1.400	1.155	1.001	0.925	0.937
RF	1.213	**0.949**	**0.887***	0.827**	0.882*

Notes: The first three rows in each panel show the root mean squared forecast error (RMSFE) for the benchmark models (AR, LBC, M-FAAR). The rest of the entries show the RMSFE of each model relative to the best performing benchmark which is the LBC model. Thus, values lower than unity suggest a particular model which utilises only Google Trends data outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon. Grey boxes denote the lowest relative error achieved by the competing models between the two subperiods. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table B.9: Google Trends Vs Macroeconomic Series: Durables

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
Panel A: Google "Related Queries"					
2017-2019					
ElNet	1.019	1.074	1.156	1.174	1.377
LASSO	1.005	1.049	1.215	1.205	1.366
adaLASSO	1.018	1.067	1.211	1.239	1.463
FAAR	1.018	0.997	1.072	1.156	1.255
CSR	1.057	1.061	1.111	1.223	1.490
RF	1.069	1.039	0.976	0.953	1.202
2020-2022					
ElNet	0.916	0.897	0.896	0.806*	0.618**
LASSO	0.877*	0.903	0.888*	0.814*	0.605**
adaLASSO	0.843*	0.903	0.862*	0.809	0.516**
FAAR	1.044	0.918	0.880*	0.844*	0.782*
CSR	0.990	0.916	0.896	0.860*	0.702*
RF	0.958	0.923	0.880*	0.818*	0.809*
Panel B: Google "Keyword Planner"					
2017-2019					
ElNet	1.032	1.119	1.168	1.186	1.065
LASSO	0.982	1.093	1.133	1.245	1.089
adaLASSO	1.010	1.050	1.203	1.235	1.183
FAAR	0.998	0.952	0.883*	0.924	0.856
CSR	1.011	1.153	1.274	1.523	1.649
RF	0.995	0.963	0.907	1.336	2.203
2020-2022					
ElNet	0.933	1.145	1.263	1.176	0.955
LASSO	0.904	0.908	0.951	0.778*	0.700*
adaLASSO	0.880*	0.955	0.973	0.879*	0.631**
FAAR	1.030	0.912	0.881	0.832*	0.844
CSR	0.995	0.935	0.918	0.869	0.765*
RF	0.924	0.881*	0.867*	0.817*	0.796*
Panel C: Google "Categories"					
2017-2019					
ElNet	0.999	0.966	0.998	0.868*	0.778*
LASSO	0.997	0.941	0.989	0.926	0.772*
adaLASSO	1.026	1.012	1.149	0.994	0.837*
FAAR	0.995	0.967	0.930	0.975	1.050
CSR	1.014	0.988	1.019	0.998	0.877
RF	1.002	0.892	0.802*	0.679**	0.976
2020-2022					
ElNet	1.035	0.963	1.003	0.800*	0.687*
LASSO	0.958	0.964	0.972	0.787*	0.675*
adaLASSO	0.911	0.997	1.002	0.767*	0.582**
FAAR	1.064	0.970	0.974	0.918	0.847
CSR	1.028	0.966	0.961	0.900	0.818*
RF	1.013	0.959	0.954	0.804*	0.780*

Notes: Entries in the table show the relative RMSFE of each model that uses only Google Trends data against the same model but with only macroeconomic predictors. Thus, values lower than unity suggest a particular model which utilises Google Trends data outperforms the corresponding macroeconomic-based model. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table B.10: Google Trends Vs Macroeconomic Series: Nondurables

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
Panel A: Google "Related Queries"					
2017-2019					
ElNet	0.974	0.940	0.996	1.062	0.995
LASSO	0.959	0.937	1.000	1.072	1.005
adaLASSO	1.006	0.935	0.997	1.104	0.970
FAAR	0.975	0.996	0.915	0.910	1.008
CSR	0.991	0.986	0.975	1.020	1.089
RF	0.984	1.013	1.005	1.039	0.946
2020-2022					
ElNet	0.944	0.964	0.956	0.919	0.704*
LASSO	0.970	0.979	0.973	0.914	0.723*
adaLASSO	0.963	0.988	0.995	0.906	0.686
FAAR	1.018	1.140	0.950	0.981	0.966
CSR	1.008	0.966	0.950	0.882	0.830*
RF	1.062	0.997	1.004	0.967	0.893
Panel B: Google "Keyword Planner"					
2017-2019					
ElNet	0.997	0.994	1.013	1.054	1.048
LASSO	0.959	0.990	1.019	0.997	1.065
adaLASSO	1.004	0.980	1.016	1.043	1.061
FAAR	1.033	1.029	0.935	0.946	0.784*
CSR	0.984	0.971	0.940	0.919	0.810
RF	0.948	1.003	0.956	0.943	0.854
2020-2022					
ElNet	0.954	0.918	0.911	0.854	0.689*
LASSO	0.973	0.913	0.926	0.845*	0.688*
adaLASSO	1.005	0.941	0.966	0.879	0.646**
FAAR	1.007	1.006	0.892*	0.904	0.836
CSR	0.984	0.922	0.905	0.817*	0.713*
RF	0.980	0.963	0.974	0.916	0.804*
Panel C: Google "Categories"					
2017-2019					
ElNet	1.019	0.983	1.035	1.087	1.004
LASSO	1.016	0.999	1.015	1.108	1.015
adaLASSO	1.000	1.018	1.024	1.110	1.010
FAAR	1.052	1.106	1.026	1.078	1.159
CSR	1.022	1.005	0.982	0.928	0.888
RF	0.963	0.993	0.972	0.933	0.803*
2020-2022					
ElNet	0.849*	0.935	0.905	0.859	0.662*
LASSO	0.906	0.937	0.909	0.845*	0.685*
adaLASSO	0.916	0.947	0.894	0.836	0.654*
FAAR	0.947	0.985	0.865*	0.832*	0.779*
CSR	0.989	0.929	0.893	0.860*	0.784*
RF	0.983	0.934	0.919	0.909	0.860

Notes: Entries in the table show the relative RMSFE of each model that uses only Google Trends data against the same model but with only macroeconomic predictors. Thus, values lower than unity suggest a particular model which utilises Google Trends data outperforms the corresponding macroeconomic-based model. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table B.11: Google Trends Vs Macroeconomic Series: Services

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
Panel A: Google "Related Queries"					
2017-2019					
ElNet	0.998	1.055	1.037	1.000	0.780*
LASSO	1.010	1.032	1.010	0.937	0.917
adaLASSO	0.999	0.999	1.003	0.925	0.714*
FAAR	1.029	1.016	1.052	0.992	0.809*
CSR	1.009	1.006	0.996	1.004	0.887
RF	1.086	1.019	1.009	0.929	0.903
2020-2022					
ElNet	0.966	1.048	1.012	1.007	0.918
LASSO	0.953	1.031	0.995	1.013	0.861*
adaLASSO	0.813*	1.043	0.997	0.979	0.826*
FAAR	1.157	1.188	1.036	1.004	1.086
CSR	1.033	0.996	0.948	0.909	0.855*
RF	0.910	0.982	1.061	1.024	0.887
Panel B: Google "Keyword Planner"					
2017-2019					
ElNet	0.965	0.956	0.965	0.904	0.878
LASSO	0.999	0.979	0.968	0.860*	0.819*
adaLASSO	0.999	0.987	1.006	0.884	0.753*
FAAR	1.032	1.019	1.031	0.944	0.879
CSR	1.030	0.965	0.982	0.971	0.853
RF	1.084	1.019	1.013	0.934	0.909
2020-2022					
ElNet	0.890	1.014	0.994	0.994	0.938
LASSO	0.932	0.980	0.994	0.995	0.895
adaLASSO	0.814*	0.980	1.010	0.967	0.882
FAAR	1.196	1.122	1.040	1.006	1.025
CSR	1.013	0.993	0.960	0.927	0.871
RF	0.933	0.977	1.024	1.036	0.902
Panel C: Google "Categories"					
2017-2019					
ElNet	1.022	1.155	1.300	1.312	1.389
LASSO	1.028	1.036	1.141	1.203	1.344
adaLASSO	1.000	1.040	1.106	1.211	1.266
FAAR	1.047	1.044	1.081	1.187	1.521
CSR	1.052	1.068	1.073	0.996	0.877
RF	1.060	1.001	0.983	0.891	0.916
2020-2022					
ElNet	0.927	1.034	1.000	0.952	0.919
LASSO	0.931	1.019	0.968	0.917	0.845
adaLASSO	0.780**	1.047	0.986	0.901	0.826*
FAAR	1.325	1.058	1.024	0.987	1.152
CSR	0.989	0.971	0.932	0.875*	0.965
RF	0.944	0.893	0.881*	0.869*	0.821*

Notes: Entries in the table show the relative RMSFE of each model that uses only Google Trends data against the same model but with only macroeconomic predictors. Thus, values lower than unity suggest a particular model which utilises Google Trends data outperforms the corresponding macroeconomic-based model. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table B.12: Google Trends Vs Macroeconomic Series: PCE

	Forecasting Horizon				
Model	h=1	h=2	h=3	h=6	h=9
Panel A: Google "Related Queries"					
2017-2019					
ElNet	0.973	0.976	0.991	0.881*	0.913
LASSO	1.005	0.973	1.034	0.886	0.779*
adaLASSO	0.999	1.006	0.996	0.850*	0.803*
FAAR	0.981	0.945	1.045	1.104	1.110
CSR	0.991	0.947	0.939	0.853*	0.783*
RF	1.063	0.987	0.914	0.894*	1.059
2020-2022					
ElNet	1.109	0.986	0.987	0.969	0.896
LASSO	1.083	1.005	1.002	0.935	0.868*
adaLASSO	0.919	1.027	0.994	0.894	0.822*
FAAR	1.089	1.096	1.020	0.977	1.010
CSR	1.037	1.013	0.966	0.904	0.870*
CSR	0.973	0.947	0.972	0.964	0.881
Panel B: Google "Keyword Planner"					
2017-2019					
ElNet	1.005	1.007	1.007	0.931	1.038
LASSO	1.021	0.972	0.989	0.883*	0.986
adaLASSO	0.999	1.005	0.950	0.950	1.021
FAAR	1.022	1.017	1.015	0.937	0.811*
CSR	0.986	0.954	0.950	0.909	0.939
RF	1.035	0.913	0.831	0.860*	0.911
2020-2022					
ElNet	1.007	0.971	0.966	0.980	0.874
LASSO	0.941	0.997	0.911	0.956	0.845*
adaLASSO	0.886*	1.011	0.922	0.931	0.790*
FAAR	1.082	1.077	1.017	0.973	0.968
CSR	1.038	0.972	0.941	0.897	0.888
RF	0.960	0.921	0.887*	0.902	0.873
Panel C: Google "Categories"					
2017-2019					
ElNet	1.041	1.076	0.992	1.049	1.276
LASSO	1.003	0.998	0.952	1.139	0.875
adaLASSO	0.999	1.021	1.016	1.100	1.038
FAAR	0.995	0.976	1.011	1.070	1.131
CSR	1.015	0.996	1.033	1.042	1.055
RF	1.048	0.980	0.904	1.043	1.258
2020-2022					
ElNet	1.025	0.966	0.989	0.906	0.912
LASSO	0.970	0.982	0.968	0.822*	0.856
adaLASSO	0.903	1.007	0.971	0.813*	0.826*
FAAR	1.102	1.103	1.055	1.028	1.140
CSR	1.044	1.021	0.964	0.912	0.916
RF	0.995	0.904	0.872	0.895	0.961

Notes: Entries in the table show the relative RMSFE of each model that uses only Google Trends data against the same model but with only macroeconomic predictors. Thus, values lower than unity suggest a particular model which utilises Google Trends data outperforms the corresponding macroeconomic-based model. The estimates were computed from 36 and 27 rolling windows covering the 2017-2019 and 2020-2022 period, respectively. Values followed by asterisks (**5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table B.13: Main Components of Personal Consumption Expenditures According to BEA's Classification System and Google Trends "Related Queries"

BEA Classification	**Google Trends**
Durable Goods	
Motor Vehicles and Parts	vehicle, motor vehicle, vehicle registration, dmv, vehicle inspection, vehicle insurance, car, car rental, car wash, used car, car insurance, rent car
Furnishings and Durable Household Equipment	furnishing, home furnishing, furniture, home furniture, home furniture stores, furnishings, home appliance, home depot, home depot appliance, lowes, lowes appliance, appliance repair
Recreational Goods and Vehicles	sports car, cars, best sports car, sports car for sale, bmw sports car, bmw
Other Durables	computer, my computer, computer science, computer screen, mac computer, laptop computer, gaming pc, pc for gaming, best gaming pc, best pc, gaming pc build, build pc
Nondurable Consumption	
Food and Beverages Purchased	food, food near, food near me, chinese food, chinese, fast food, beverages, alcoholic beverages, alcoholic, alcohol, alcohol beverages, beverage
Clothing and Footwear	clothing, clothing stores, clothing stores, clothes, mens clothing, women clothing
Gasoline and Other Energy Goods	gasoline, gas, what is gasoline, gasoline price, gasoline prices, gasoline can, oil, oil change, oil filter, essential oil, oil prices, oil price
Other Nondurable Goods	pharmacy, cvs pharmacy, cvs, walmart pharmacy, pharmacy hours, walgreens pharmacy
Services Consumption	
Housing and Utilities	electricity, energy, electric, electricity cost, electricity bill, solar electricity
Health Care	health, health care, health insurance, health department, mental health, public health
Transportation Services	transportation, transportation department, department of transportation, public transportation, airport transportation, bus transportation
Recreation Services	gambling, online gambling, casino, casino gambling, sports gambling, gambling age, museum, art museums, free museums, museums near me
Food Services & Accommodation	hotel, hotels, las vegas hotel, grand hotel, hilton hotel, hilton, restaurant, mexican restaurant, restaurant near me, chinese restaurants, restaurants
Financial Services and Insurance	insurance, health insurance, car insurance, life insurance, auto insurance, home insurance
Other Services	package delivery, package, delivery, delivery package, ups, ups package delivery

Table B.14: Main Components of Personal Consumption Expenditures According to BEA's Classification System and Google Trends "Keyword Planner"

BEA Classification	**Google Trends**
Durable Consumption	
Motor Vehicles and Parts	car, auto, car dealerships, classic cars for sale, car leasing, car auctions, vehicle, cars for sale, used cars, used cars for sale, honda crv, ford f150
Furnishings and Durable Household Equipment	furnishing, modern furniture, home decor stores, furniture style, vintage furniture, rustic furniture, home appliance, air purifier, appliances, kitchen appliances, water dispensers, washing machine
Recreational Goods and Vehicles	sports car, cheap sports cars, best sports cars, sports cars for sale, race car driving, sports suv
Other Durables	computer, pc, monitor, gaming pc, laptops, desktop computer, pc builder, gaming computer
Nondurable Consumption	
Food and Beverages Purchased	food, chilis, olive garden, taco bell menu, beverage, lemonade, long island iced tea, drinks, smirnoff ice, coca cola coke
Clothing and Footwear	clothing, jeans, sweater, sweatshirt, jacket, shorts
Gasoline and Other Energy Goods	gasoline, gas stations, gas prices, shell gas stations, chevron gas station, shell gas cards, oil, oil price, crude oil price, change oil, castor, valvoline coupon
Other Nondurable Goods	pharmacy, drugstore, online pharmacy, 24 hour pharmacy, walgreens pharmacy, cvspharmacy
Services Consumption	
Housing and Utilities	electricity, electric, electric company, service electric, power to choose, electrification
Health Care	health, health department, hipaa, epidemiology, healthy food, health care
Transportation Services	transportation, freight, department of transport, transportation services, car transport, public transport
Recreation Services	gambling, bets, online betting, online casinos, betting online, online gambling, museum, natural history museum, art museum, smithsonian museums
Food Services & Accommodation	hotel, accommodation, resort, mgm grand, wynn las vegas, borgata, restaurant, mexican restaurant
Financial Services and Insurance	insurance, car insurance, life insurance, medicaid, car insurance quotes, cheap car insurance
Other Services	package delivery, ups shipping, cheapest overnight shipping, ups delivery, fedex delivery

Table B.15: Main Components of Personal Consumption Expenditures According to BEA's Classification System and Google Trends "Categories"

BEA Classification	**Google Trends**
Durable Consumption	
Motor Vehicles and Parts	Autos & Vehicles, Commercial Vehicles, Classic Vehicles, Vehicle Shopping, Vehicle Parts & Accessories, Vehicle Maintenance, Motorcycles, Motor Vehicle, Vehicle, Car Dealership, Car Electronics, Vehicle Wheels & Tires
Furnishings and Durable Household Equipment	Home Appliances, Home Furnishings, Home Improvement, Homemaking & Interior Decor, Major Kitchen Appliances, Home Financing, Bed & Bath, Garden Furniture, Home Bar, Cookware & Diningware
Recreational Goods and Vehicles	Movies, Music Equipment & Technology, Camera & Photo Equipment, Books & Literature, Computer & Video Games, Recreational Aviation, Motor Sports, Sports Car, Recreational Vehicle, Television Set, Musical Instrument, Desktop Computers, Sporting Goods, Audio Equipment
Other Durables	Computer Hardware, Consumer Electronics, Eyeglasses & Contacts, Communications Equipment
Nondurable Consumption	
Food and Beverages Purchased for Off-Premises Consumption	Alcoholic Beverages, Non-Alcoholic Beverages, Alcoholic Drink, Food & Drink, Food Production, Food Service, Grocery & Food Retailers
Clothing and Footwear	Apparel, Footwear, Clothing Industry, Clothing, Athletic Apparel, Casual Apparel
Gasoline and Other Energy Goods	Electricity, Energy & Utilities, Oil & Gas, Gasoline, Fuel
Other Nondurable Goods	Pharmaceutical Drug, Toys, Animal Products & Services, Pets, Tobacco Products, Hair Care, Oral & Dental Care, Beauty & Fitness, Newspapers, Magazines, Pharmacy, Cleaning Supplies & Services, Stationery, Linens, Pet Store, Face & Body Care, Health
Services Consumption	
Housing and Utilities	Housing, Home Financing, Electricity, Natural Gas, Water Supply, Sanitation
Health Care	Oral & Dental Care, Pediatrics, Medical Facilities & Services, Health, Nursing, Hospital
Transportation Services	Vehicle Maintenance, Automobile Repair Shop, Motor Vehicle Service, Bus & Rail, Air Travel, Car Rental & Taxi Services, Travel, Cruises & Charters
Recreation Services	Gambling, Casino, Veterinarians, Animal Products & Services, Regional Parks & Gardens, Theme Parks, Libraries & Museums
Food Services & Accommodation	Restaurants, Food & Drink, Lodging, Hotels & Accommodations, Grocery & Food Retailers
Financial Services and Insurance	Insurance, Health Insurance, Home Insurance, Financial Services, Investing, Credit & Lending
Other Services	Education, Face & Body Care, Internet & Telecom, Mail & Package Delivery, Social Services

Chapter 4

Forecasting Inflation Uncertainty with Internet Search and News Predictors[1]

4.1 Introduction

Accurately predicting inflation uncertainty in the near and medium term carries significant implications for economic policy and broader business decisions within the economy. Mishkin (2008) argues that policymakers should assess the costs of inflation in terms of both its level and its uncertainty, while Giordani and Söderlind (2003) states that relying solely on point forecasts of inflation is adequate only under limited conditions. Shifts in inflation uncertainty can provide valuable signals regarding the credibility of policy actions. Credible inflation targeting policies imply stable and predictable inflation. Therefore, rising inflation uncertainty may indicate unmoored underlying inflation expectations and growing risks that actual inflation could persistently deviate from the central bank's objective.

Additionally, recents events such as the Covid-19 pandemic, supply constraints in energy markets, and rising geopolitical tensions have amplified economic uncertainty and highlighted the sensitivity of inflation to disruptions in social mobility, international trade, supply chains, and production. Therefore, given the current surge in uncertainty around

[1]Co-authored with Michael P. Clements, Andrew Urquhart, and Oguzhan Cepni.

inflation, it is more critical than ever to effectively monitor and forecast inflation uncertainty. Moreover, the decisions of economic agents are influenced by their evaluation of the probabilistic distribution of predicted economic data meaning that higher moments are of primary importance. Timely and accurate predictions of inflation uncertainty is crucial considering that increasing uncertainty surrounding inflation could disrupt economic stability, delay investment plans, weigh on consumer confidence and affect saving and spending patterns.

In recent years, technological advancements and our capacity to store vast amounts of data have led to the emergence of new data sources. These sources have the potential to offer real-time insights into consumer behavior, business activities, and other economic variables that official statistics may not capture in a timely manner. These alternative data sources are typically formed as the outcome of business transactions, social media activity, news articles, or internet searches, among others. The frequency of internet searches, as reported by Google Trends, can provide valuable insights into the current and future dynamics of the economy. During periods marked by heightened uncertainty, individuals often exhibit a tendency to augment their information-seeking behaviour to better understand prevailing economic conditions. Furthermore, the role of news in today's society as the primary source of information is indisputable. Economic agents rely on news to form expectations about future dynamics. News articles often contain narratives associated with the current conditions in the economy or economic policy decisions. Therefore, it is reasonable to assume that the frequency of specific keywords appearing in news articles conveys vital information regarding the current and future state of the macroeconomy.

That being said, the primary objective of this empirical chapter is to uncover the predictive power of alternative data sources in forecasting inflation uncertainty in the United States, using data spanning from January 2004 to December 2022. Considering that forecasting the statistical properties and the behaviour of inflation uncertainty presents an empirical challenge due to the lack of an objective measure of uncertainty, we construct three uncertainty indexes each one aiming to capture the uncertainty produced by a different group of economic agents, that is, households, professional forecasters, and investors. Two main sources of alternative predictors are employed to forecast the inflation uncertainty indexes: Google Trends and News Trends. Google Trends provides information about online search volume on specific queries, while News Trends, available from Bloomberg, aggregates keywords by topic over time and measures the frequency of these terms appearing in news articles. To construct a comprehensive set of terms related to inflation uncertainty, we start from a primitive keyword, which in our case is "inflation". Subsequently, we utilise the "related-keywords" function provided by Google to expand the dataset with semantically related queries, thereby forming the set of predictors.

To assess the predictive capabilities of these two alternative sources of data, we conduct a monthly out-of-sample forecasting exercise and compare the predictive accuracy in forecasting inflation uncertainty generated by Google and News Trends against a set of macroeconomic factors. Given that some of our forecast experiments involve over 35 predictors, and conventional estimation techniques tend to exhibit a poor forecasting performance when dealing with a large number of covariates, we rely on methods suitable for data-rich structures. Thus, our approach closely follows the methodological procedure of Borup and Schütte (2022) and uses both linear and non-linear models such as bagging, complete subset regressions (CSR), and random forests. Notably, previous studies have highlighted the effectiveness of these methods. For instance, Garcia et al. (2017) document the superiority of complete subset regressions in forecasting inflation in Brazil, while Kotchoni et al. (2019) show that CSR exhibits sufficient forecasting power when predicting macroeconomic variables. On a similar note, Medeiros et al. (2021) provides evidence that random forests yields the most accurate U.S. inflation forecasts, while Rapach and Strauss (2012) imply that bagging can perform quite satisfactorily when forecasting employment growth.

Internet search data have been extensively used by the empirical literature to forecast a wide range of macroeconomic variables. Specifically, researchers have explored the usefulness of Google Trends data in forecasting output growth (Woloszko, 2020; Ferrara and Simoni, 2023; Bantis et al., 2023), consumption (Vosen and Schmidt, 2011; Woo and

Owen, 2019), housing prices (Møller et al., 2023), labour market developments (D'Amuri and Marcucci, 2017; Borup and Schütte, 2022), and price levels (Seabold and Coppola, 2015). Overall, these studies demonstrate that Google Trends has the capacity to yield substantial forecasting improvements at different time horizons, contingent upon the specific target variable and the chosen methodological framework.

Google search data may contain valuable insights into consumer and other agents' plans and intentions as well as reflect sentiment and concerns about economic fundamentals and financial markets. Several studies in economic psychology indicate that individuals tend to increase their online searches when faced with greater price uncertainty (Lemieux and Peterson, 2011; Abbas et al., 2013). Therefore, we can assume that Google searches reflect, to some extent, the concerns of individual economic agents and therefore can potentially track quite well inflation uncertainty indicators.

Moreover, recently there is a growing amount of studies that attempt to measure and forecast sentiment, uncertainty, and other macroeconomic indicators using textual analysis. In particular, Baker et al. (2016) construct an economic and political uncertainty index using news articles that contain keywords of interest; Shapiro et al. (2022) derive a time-series indicator of economic sentiment based on economic and financial newspaper articles, Kalamara et al. (2020) attempt to extract signals using newspaper textual data to forecast various macroeconomic variables, Thorsrud (2020) utilises business newspaper data to construct daily business cycle index, while Barbaglia et al. (2023) construct sentiment measures based on news articles to forecast a range of key macroeconomic variables. These papers, in overall, document that news articles provide informational content that is relevant to several applications and therefore can be a valuable source of data for researchers.

While our approach does not incorporate textual analysis, we make use of Bloomberg's News Trends indicators to capture the levels of uncertainty surrounding inflation. The typical response during uncertain times is to generate more news content about the evolution of key macroeconomic data and therefore extracting the frequency with which certain queries appear on articles could be an insightful source of information.

Google and News Trends offer distinct advantages over traditional variables commonly employed in macroeconomic forecasting. Official statistics typically exhibit a considerable publication lag and are subject to substantial revisions. Soft data like household and business surveys, while potentially more timely and less prone to revisions, could entail significant expenses in their acquisition and may be susceptible to selection biases. Conversely, Google and News Trends provides the advantage of real-time availability as well as the ability to focus on specific geographical regions. Moreover, the simplicity and cost effective way of obtaining additional Google Trends series makes it easy for researchers to expand their set of predictors.

The contribution of this chapter is two-fold. First we contribute to the growing literature of monitoring and forecasting uncertainty using alternative high-frequency datasets. Several papers have employed Google Trends data to track developments around various uncertainty indexes. For instance, Castelnuovo and Tran (2017) construct internet-search indexes to track business cycle uncertainty using keywords that appear frequently in monetary policy official documents and Bilgin et al. (2019) utilise Google Trends data to create an economic and financial uncertainty index for Turkey. Also, Dzielinski (2012) creates a search-based uncertainty index to measure investors' confidence and uncertainty, while Donadelli (2015) employ Google search data to measure policy-related uncertainty. However, according to our knowledge, this is the first study that links Google and News Trends datasets to forecast inflation uncertainty.

Second, we contribute to the growing field of forecasting inflation dynamics with a large set of predictors using high-dimensional and machine learning models. Garcia et al. (2017) employ a large number of machine learning models to forecast inflation in Brazil; Kohlscheen (2022) attempts to predict inflation across 20 advanced economies using random forest; Cheng et al. (2021) utilise machine learning methods to aggregate survey-based individual information and to forecast inflation in the United States; while Medeiros et al. (2021) uses a large dataset to forecast U.S. inflation using a wide range of machine learning models and methods suitable for high-dimensional datasets. Therefore, this study aims to contribute to this area by shedding light on the predictive power of high-dimensional and machine learning models for forecasting an uncertainty indicator.

Several interesting takeaways are extracted from our empirical analysis. First, when forecasting households' inflation uncertainty, Google Trends predictors provide striking predictability at horizons between one and three months ahead with specifications based on bagging and complete subset regressions exhibiting the best forecasting performance. While they outperform significantly models that relying on macroeconomic predictors at short horizon forecasts, their accuracy declines as the prediction horizon widens and tends to converge to levels similar to the ones obtained using macroeconomic factors. News Trends perform also satisfactorily and in some cases the combination of Google and News Trends provides the highest forecast accuracy. However, when forecasting investors' inflation uncertainty, the macroeconomic factors yield the most accurate predictions at all horizons except $h = 1$. Forecast errors for professional forecasters' inflation uncertainty are broadly similar between the alternative and traditional predictors, with macroeconomic factors performing slightly better at short forecast horizons while Google Trends provide some gains at $h = \{9, 12\}$ horizons.

Second, the bagging specifications deliver in most cases the most accurate predictions. Except when the target variable is based on consumer surveys where both bagging and CSR models provide the lowest forecast errors compared to the benchmark, bagging exhibits in overall the best forecasting performance.

Third, investigating predictability over time suggests that when forecasting households' uncertainty, the macroeconomic factors appear to be the most informative source of data until 2017-2020 while Google and News Trends provide significant gains during the Covid-19 pandemic and period of increased uncertainty due to geopolitical tensions and supply shocks. Nevertheless, when forecasting investors' uncertainty, the macroeconomic factors consistently provide the highest accuracy over the out-of-sample period for horizons between three and one year ahead.

Finally, in an attempt to shed light on which Google Trends series contain the highest predictive power, when the target variable is based on consumer surveys the Google Trends series that were selected as the most informative centered predominantly around queries related to "inflation", while when the uncertainty measure is based on market indicators and professional forecasters the selected queries include additional terms related to the "economy" and "interest rates".

The rest of this chapter is organised as follows: Section 4.2 describes the construction of inflation uncertainty measures and the Google and News Trends predictors, and Section 4.3 discusses the methodological framework. Section 4.4 presents the forecasting setup and empirical findings, while Section 4.5 summarises the main empirical results of this chapter. Finally, Appendix C contains additional findings including a robustness analysis.

4.2 Data Description

This section describes in detail the selection of the indexes that are used as a proxy for inflation uncertainty as well as the construction of the alternative groups of predictors utilised to forecast inflation uncertainty. In particular, Section 4.2.1 discusses the three measures of inflation uncertainty that have been chosen and the rationale behind these selections, while Section 4.2.2 outlines the construction of the Google and News Trends datasets as well as the selection of the macroeconomic data used as a benchmark in our forecasting experiments.

4.2.1 Measuring Inflation Uncertainty

Considering that inflation uncertainty is intrinsically unobservable, assessing its estimates poses challenging methodological issues. Measuring inflation uncertainty requires a careful thought since a one-size-fits all answer does not exist. Therefore, we utilise several disagreement indicators based on household, market participants and professional forecasters expectations. In particular, the following inflation uncertainty indicators are used in this study: (i) the interquartile range of expected price changes during the next year and the next five years extracted from the University of Michigan Surveys of Consumers, (ii) the interquartile range of inflation point forecasts derived from Bloomberg's Individual Economist Estimates, and (iii) the 12-month rolling standard deviation of breakeven inflation rates based on five-year, seven-year, and 10-year instruments.

The motivation behind the construction of simple disagreement-based uncertainty measures is based on the fact that disagreement indicators derived from survey and market expectations leverage forward-looking information, rendering them particularly suitable to reflect the properties of uncertainty proxies (Claveria, 2021). Moreover, Dovern (2015) finds a positive correlation between forecaster disagreement and realised stock market volatility as well as uncertainty indicators derived from newspaper articles. Additionally, Giordani and Söderlind (2003) provide empirical evidence that there is a significant correlation between disagreement and uncertainty, concluding that the former serves as a suitable proxy variable.

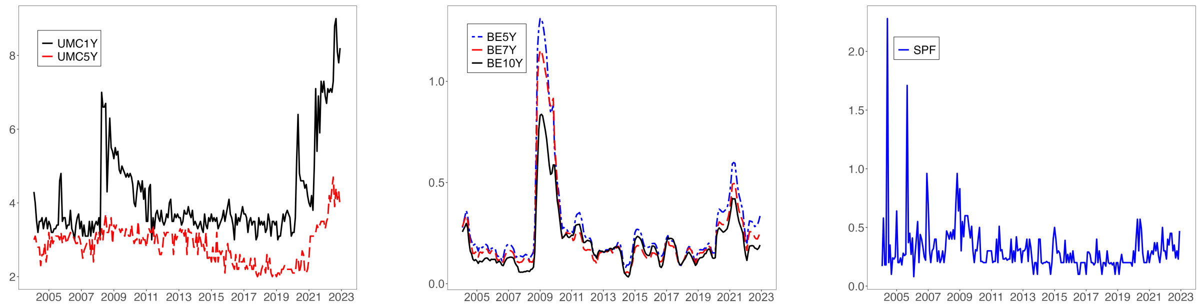
The first measure of inflation uncertainty utilised in this chapter attempts to capture uncertainty around households' expectations about price developments. Consumer surveys provide a direct measure of consumers' inflation expectations and are based on a large-scale survey among individuals. We take the view that a larger interquartile range in households' responses reflect a greater uncertainty about future inflation dynamics. In a similar manner, Baker et al. (2016) use the interquartile range of quarterly individual inflation forecasts to construct a dispersion component as a proxy for uncertainty. It should be noted, however, that consumer survey-based measures have been criticised since in some cases consumers have low economic incentives to report their expectations accurately (Keane and Runkle, 1990) and also because in some cases there might be issues regarding the representative of the survey sample (Manski, 2004).

The second measure for inflation uncertainty employed in this study is derived from professional forecasters. It is believed that the appointment procedure of the respondents and the fact that some of the forecasters are close to important policymakers ensures that their responses describe reasonably well their expectations (Giordani and Söderlind, 2003). Additionally, professional forecasters should be better informed and better able to answer to technical questions (Clements et al., 2023). Similarly to consumers' price expectations, we take the interquartile range as an indicator for uncertainty.[2] Heuristically,

[2]Our initial choice was to follow Giordani and Söderlind (2003) and measure the dispersion of professional forecasters' point predictions by computing the difference between the 84th and 16th percentiles of

people may disagree more when greater uncertainty prevails.

The third measure of inflation uncertainty is a market-based indicator that relies on breakeven inflation rolling standard deviation. The breakeven inflation rate represents the market's expectation for the average annual inflation over a specified horizon and is computed as the difference between the yield to maturity on nominal bond and inflation-linked bond with the same maturities. When market participants anticipate higher inflation rate, they typically seek for higher nominal yields causing the breakeven inflation rate to rise. Therefore, it is reasonable to assume that during periods of heightened uncertainty, the fluctuation of the breakeven inflation rate tends to rise, leading to higher rolling standard deviation estimates, reflecting investors' uncertainty about the path of future inflation. On a similar note, Altig et al. (2020) use the standard deviation of one-year ahead GDP growth forecasts as a proxy for uncertainty.



(a) Households' Uncertainty **(b)** Investors' Uncertainty **(c)** Forecasters' Uncertainty

Figure 4.1: U.S. Inflation Uncertainty Indexes

Notes: This graph shows the U.S. inflation uncertainty indexes. The plot on the left-hand side shows consumer survey-based measures of inflation uncertainty, the plot in the middle panel shows the market-based measures of inflation uncertainty, while the plot on the right-hand side exhibits the inflation uncertainty index based on the disagreement between professional forecasters.

Figure 4.1 shows the measures of U.S. inflation uncertainty employed in this chapter. The graph clearly shows that all series exhibit a spike during the Global Financial Crisis of 2007-2009, with this trend being most pronounced in the market-based indicators.

the point forecasts. However, considering that a similar approach is not feasible in the case of households' uncertainty due to the lack of necessary information to construct such an index, we take the interquartile range for consistency purposes.

While all series experienced a sharp increase during the early phase of the Covid-19 pandemic, the market-based measures and the index derived from the disagreement among professional forecasters display a mean-reversion behaviour. In contrast, the consumer-based indicators present a steep upward trend without signs of correction, at least until the end of our sample period. In addition, the index based on professional forecasters is more volatile, exhibiting several abrupt spikes, especially during the first part of the sample.

4.2.2 Constructing the Google and News Trends Datasets

The main goal of this chapter is to investigate whether alternative sources of information could provide forecasting benefits for measures of inflation uncertainty. The two main sources of alternative predictors used in our analysis are Google Trends and Bloomberg's News Trends. Google Trends provides information about the frequency with which specific keywords, within a particular geographic area, are searched by users relative to the overall search volume. Thus, it is reasonable to assume that internet search activity for phrases like "what is inflation" or "current inflation rate" likely reflects genuine interest in learning more about current price developments. This interest typically increases when there are growing concerns about inflation and its effects on real purchasing power and disposable incomes.

The Google Trends series are provided in various frequencies (e.g., daily, weekly, etc.) and their availability starts from January 2004. Google assigns an index to each search query ranging from zero to 100. A value of zero reflects the lowest relative user interest or insufficient data for the query, while a value of 100 signifies the highest point of popularity. An appealing feature of Google Trends for forecasters is that they can be acquired in a simple and costless way without any publication delay. In contrast, traditional macroeconomic data are typically available several weeks after the end of the reference month and, in some cases, are prone to considerable revisions. Additionally, while several consumer and business surveys are available in a timely manner, the cost of obtaining them can be significant.

Bloomberg's News Trends, the second source of alternative predictors in our forecasting experiments, provides a quantitative measure of media attention for specific keywords. Utilising more than 150,000 global authoritative sources and advanced artificial intelligence models, News Trends quantifies the frequency with which particular keywords appear in news articles, reflecting essentially the real-time interest of the media in specific terms. Therefore, it is plausible to expect that increased news coverage of inflation-related topics might contain valuable information for tracking and predicting developments in inflation uncertainty.

The starting point for constructing both groups of predictors is the selection of a primitive keyword, which in our case is "inflation". Next, we utilise Google Trends' related-keywords function to derive a list of 25 keywords that are semantically related to the primitive query. This list of related queries, presented in Table 4.1, is then used to construct the Google Trends and Bloomberg's News Trends groups of predictors. However, some series from News Trends exhibit low volume and have therefore been excluded from our analysis. Only the terms in bold have remained in this group of predictors. Our sample period spans from January 2004 to December 2022 and uses a monthly frequency. Figure 4.2 displays two selected predictors, namely "Inflation" and "Interest Rates", extracted from both Google Trends and News Trends for the United States. The series exhibit a considerable upward trajectory during the Global Financial Crisis, as well as a steep upward trend over the Covid-19 pandemic and the recent inflationary episode, indicating their potential predictive power when forecasting an uncertainty index.

To compare the forecast gains obtained from the two alternative group of predictors, a set of macroeconomic predictors is employed derived from the McCracken and Ng (2016) database. Given that the cross-sectional dimension of this database differs significantly from that of Google Trends and News Trends – as it includes more than 120 variables – we use the first eight principal components as a means of reducing the dimensionality of this dataset.

Table 4.1: Google Related-Keywords Generated List using the Keyword "inflation"

inflation	**cpi inflation**
inflation calculator	inflation index
what is inflation	**interest rate**
rate of inflation	economic inflation
us inflation	current inflation rate
inflation rates	calculate inflation
inflation definition	dollar inflation
gdp	definition of inflation
economy	what is the inflation rate
what is inflation rate	**unemployment rate**
us inflation rate	**cpi**
interest rates	**inflation rate**
current inflation	

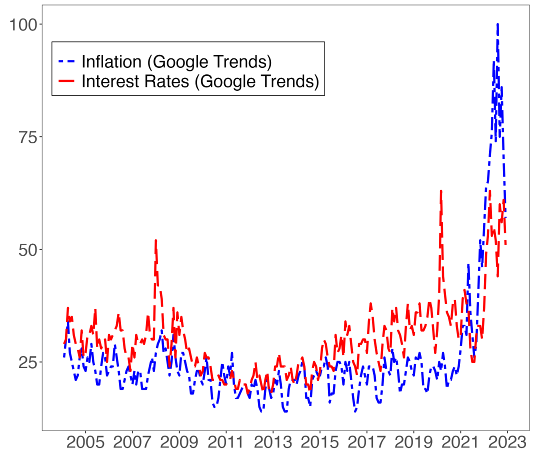
Notes: The table presents the list of keywords generated by Google Trends' related-keywords function, using the primitive keyword "inflation". All of these keywords are included in the Google Trends dataset, while for the News Trends dataset, only the terms in bold are chosen. The remaining series have been excluded from our analysis due to their low volume, as they are deemed unsuitable for any use in our econometric models.

4.3 Methodological Framework

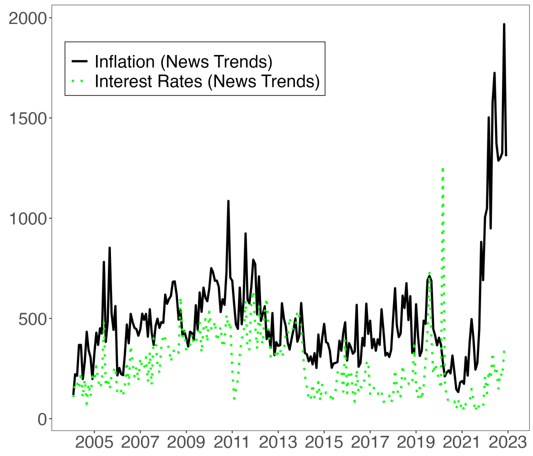
In this section, we present our methodological framework and provide a brief overview of the techniques employed to make predictions and evaluate their performance. Section 4.3.1 describes the bagging technique, Section 4.3.2 the complete subset regression while Section 4.3.3 outlines the random forest model.

4.3.1 Bagging

Bagging, introduced by Breiman (1996) and stands for *bootstrap aggregation*, is a statistical technique aimed at reducing the mean-squared error of the out-of-sample predictions by combining forecasts from several unstable models estimated over different bootstraps subsamples. In the context of implementing the bagging model, we adopt the approach of Inoue and Kilian (2008) where the first step consists of the estimation of an unrestricted model that incorporates all potential covariates for each bootstrap sample and



(a) Google Trends



(b) News Trends

Figure 4.2: U.S. Google and News Trends "Inflation" and "Interest Rates"

Notes: This graph shows the Google and News Trends "Inflation" and "Interest Rates" series extracted for the United States.

then selects those predictors with a corresponding absolute t-statistic that is greater than a pre-specified critical value c:

$$\hat{Y}_{j,t+h} = \hat{\alpha}_i + \sum_{i=1}^{N} \hat{\beta}_i X_{i,t}^{P*}$$
$$X_{i,t}^{P*} = \begin{cases} X_{i,t}^{P}, & \text{if } \left|t_{X_i}\right| > c \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

where $\hat{Y}_{j,t+h}$ denotes the jth target variable at forecast horizon h, $X_{i,t}^{P*}$ reflects the selected ith individual predictor that belongs to the X_t^P group of predictors with $P = \{MF, GT, NT, GTNT\}$, and t_{X_i} is the t-statistic computed on the slope parameter $\hat{\beta}_i$. Thus, from the predictors included in X_t^P we select only those that are statistically significant. The t-statistics are constructed using the Newey et al. (1987) robust standard errors with lag truncation equal to $h-1$. We follow Borup and Schütte (2022) and we set the critical value equal to 2.58 and thus only the predictors that are significant at the 1% level are selected.[3]

[3]For robustness we also used a critical value equal to 1.96 corresponding to statistical significance at the 5% level. Findings, which can be found in Table C.3, do not show material differences compared to the main empirical results presented in Section 4.4

Bagging can be implemented for the pretesting procedure using a moving-block bootstrap, a method used to effectively reduce the variance arising from model uncertainty. We produce $B = 800$ bootstrap samples for the left- and right-hand-side variables in Equation (4.1) by randomly selecting blocks of size m (with replacement). For each bootstrap sample we compute Equation (4.1) using information only up to time t and then the bootstrap forecast, $\hat{Y}_{b,j,t+h}$, is estimated using the bootstrap coefficients and original data $X_{i,t}^{P^*}$. The final prediction is calculated as the average of the B bootstrap forecasts:

$$\hat{Y}_{j,t+h} = \frac{1}{B}\sum_{b=1}^{B}\hat{Y}_{b,j,t+h} \tag{4.2}$$

By employing resample averaging, bagging effectively mitigates the instability inherent in the decision rule implying that the variance of the bagged model is expected to be less than that of the model selected solely based on the original data. The potential autocorrelation in $\hat{Y}_{j,t+h}$ is controlled by implementing the circular block bootstrap of Politis and Romano (1991) with a block size determined according to Politis and White (2004).

4.3.2 Complete Subset Regressions

The underlying principle of complete subset regression (CSR), introduced by Elliott et al. (2013), is to identify the optimal subset of $X_t^{P^{**}} \in X_t^P$ variables for predicting $Y_{j,t+h}$ by evaluating all possible combinations of predictors. However, due to the potential computational burden, CSR addresses this challenge by fixing the number of explanatory variables and generating forecast estimates through the aggregation of all feasible linear regression models. When we have N^P regressors that correspond to the Pth group of predictors, CSR selects $q \leq N$ predictors and estimates all possible q-dimensional combinations from N^P. Multiple predictions of the target variable are generated by considering different combinations of subsets of predictor variables X^P. To arrive at the final forecast, all these individual predictions are combined using a simple averaging technique:

$$\begin{aligned} \hat{Y}_{j,l,t+h} &= \hat{\alpha}_l + \hat{\beta}_l X_{l,t}^P \\ \hat{Y}_{j,t+h} &= \frac{1}{M}\sum_{l=1}^{M} \hat{Y}_{j,l,t+h} \end{aligned} \tag{4.3}$$

where $X_{l,t}^P$ denotes the q-vector of predictors that belong to the Pth group for each model $l = 1, \ldots, M$. We follow Borup and Schütte (2022) and choose a parsimonious specification with $q = 3$ predictors.[4]

4.3.3 Random Forests

Although the concept of capturing nonlinearities through decision trees is intuitive and attractive, primarily due to its potential for interpretability, the resulting predictions often suffer from high variance. The process of fitting recursive trees suffers from two main issues: (i) instability and (ii) a tendency to overfit. However, a more effective and widely adopted solution was proposed by Breiman (2001), namely, random forests. This approach involves growing multiple trees on subsamples of the data, introducing further randomisation by considering only a subset of predictors for each potential split. The key hyperparameter to be determined is the number of variables to be evaluated at each split. The predictions of the individual regression trees are then averaged to form a single "ensemble" prediction for the target variable.

If we denote by R_z the partition of covariates space with $z = 1, \ldots, Z$ and Z denoting the number of terminal nodes and g_Z the node means, then the forecast for the j-th uncertainty index given a set of predictors X_t^P based on regression trees is computed as follows:

$$Y_{j,t+h} = \sum_{z=1}^{Z} g_z I_{z(X_t^P, \theta_z)} \tag{4.4}$$

where Z is the number of terminal nodes, θ_z is a set of parameters that define the z-th partitioning region, and $I_{z(X_t^P, \theta_z)}$ is an indicator function that is equal to one when $X_t^P \in R_z(\theta_z)$. Random forests is a collection of regression trees, each estimated based on

[4] We also run forecast experimentations with $q = \{4, 5\}$ but results, as shown in Table C.1 and Table C.2, remain broadly similar with the ones presented in Section 4.4.

a bootstrap sample $b = 1, \dots, B$ of the original predictors. For each bootstrap sample b, a tree with Z_b regions is estimated for a randomly selected subset of the predictors. The final forecast is computed as the average of predictions produced by each regression tree:

$$\hat{Y}_{j,t+h} = \frac{1}{B}\sum_{b=1}^{B}\Big[\sum_{z=1}^{Z_b}\hat{g}_{z,b}I_{z,b(X_t^P,\hat{\theta}_{z,b})}\Big] \tag{4.5}$$

The minimum number of leaf node observations is specified to five regression trees. The number of predictors randomly selected for each decision split is equal to 1/3 of the number of variables for regression trees. As in bagging, the number of bootstrap samples is set to 800.[5]

4.4 Empirical Results

4.4.1 Setup and Main Forecast Exercise

In this section, we present the out-of-sample forecasting performance of our novel Google and News Trends datasets using various model specifications. The sample period is from January 2004 to December 2022 and all competing models are estimated using a rolling window scheme with 48 observations. Hence, the initial forecasts for horizons up to six months are made during the early months of 2008. Considering that the pandemic period is also included in our sample, this provides us with a unique opportunity to evaluate the forecast benefits of various competing model specifications during both the Global Financial Crisis of 2007-2009 and the recent economic downturn caused by the Covid-19 pandemic.

Our main interest is to evaluate the predictive capabilities of alternative sources of data such as Google Trends and News Trends in comparison to a set of macroeconomic factors and a simple benchmark model. Regarding the choice of the benchmark, we have opted for the rolling mean forecast, which serves as a natural reference model since our models encompass it. In other words, if the model in question performs better than the

[5] Forecast experiments with bootstrap samples equal to 700 and 900 were also estimated with findings indicating stable forecast accuracy.

simple rolling mean model, it implies that the corresponding model adds value beyond what can be achieved by just using the rolling average of the past observations as a forecast.

The statistical significance of the forecasting results is evaluated using the Diebold-Mariano (DM) test of equal predictive accuracy (Diebold and Mariano, 1995). The DM test is employed under the null hypothesis that the model being examined does not outperform the benchmark, while the alternative hypothesis suggests that it does. To evaluate the forecast accuracy of each model the root mean squared forecast error (RMSFE) is computed as follows:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_{j,t+h} - \hat{Y}_{j,t+h})^2} \tag{4.6}$$

where $Y_{j,t+h}$ and $\hat{Y}_{j,t+h}$ are the actual and predicted values of the jth target variable at the h forecast horizon, respectively, while T denotes the total number of predictions. In addition, the Campbell and Thompson (2008) out-of-sample R^2 is also used to inference predictability in our empirical exercise. The R^2_{OoS} is defined as follows:

$$R^2_{OoS} = 1 - \frac{\sum_t (Y_{j,t+h} - \hat{Y}_{j,m,t+h})^2}{\sum_t (Y_{j,t+h} - \bar{Y}_{j,t+h})^2} \tag{4.7}$$

where $Y_{j,t+h}$ is the realisation of the jth target variable, $\hat{Y}_{j,m,t+h}$ represents the forecast for the jth target variable generated by the $m = \{Bagging, CSR, RFs\}$ model specification at the h forecast horizon, while $\bar{Y}_{j,t+h}$ denotes the rolling mean prediction. The R^2_{OoS} lies in the interval $(-\infty, 1]$ with negative numbers implying that the benchmark outperforms the m model specification.

All the series employed in the forecasting experiments have been transformed into a stationary form. In order to facilitate this process and considering the fact that some series appear to contain unit roots while others appear to be stationary, we employ a systematic testing approach based on Ayat and Burridge (2000). This approach involves a sequential testing strategy by testing successively for stationarity, linear trend stationarity, and quadratic trend stationarity using an Augmented Dickey-Fuller (ADF) test. If the series contain deterministic trends, the trend is removed. If they have a unit root, then first-differences are employed. We utilise the Bayesian Information Criterion (BIC) to determine the lag length and use a 1% level of significance. If the series are found to contain a seasonal element, it is removed by regressing each series on monthly dummies and extracting the residuals of this regression. To avoid "look-ahead" bias and to align the data pre-processing with an out-of-sample exercise we implement the sequential testing for unit roots on a recursively expanding window where the initial window is set equal to 48 observations in line with the estimation window used in our forecasting experiments.[6]

Tables 4.2 - 4.4 summarise the results of our main prediction experiments, in which we evaluate the forecasting performance of several model specifications (bagging, CSR, random forests) against the rolling mean benchmark for three different measures of inflation uncertainty, that is, consumer-based surveys, market-based indicators, and surveys from professional forecasters. In particular, Table 4.2 presents the main empirical findings when forecasting the interquartile range of one- and five-year consumer price expectations (UMC1Y, UMC5Y), Table 4.3 when forecasting the five- and seven-year breakeven inflation rates, while Table 4.4 when the target variable is the 10-year breakeven inflation rate (BE10Y) as well as the disagreement between professional forecasters defined by taking the interquartile range of monthly individual inflation forecasts submitted by economists on Bloomberg.

Each panel is separated into three sections. The first contains the results of the bagging-based models, the second shows the results of the CSR model specifications, while the third exhibits the findings of random forest models. Each specification utilises four set of predictors, namely, macroeconomic factors (MF), News Trends (NT), Google Trends (GT), and Google Trends and News Trends combined (GTNT). All entries in the table represent relative RMSFEs, in which the numerator is the forecast error of each corresponding model while the denominator reflects the forecast error of the benchmark model. Therefore, values below unity imply that the model of reference outperforms

[6]The term "look-ahead bias" refers to our attempt to avoid using information that would not have been available in real-time as this could potentially distort the forecast accuracy of our experiments.

the rolling mean benchmark model. Empirical results are reported at monthly forecast horizons of $h = \{1, 3, 6, 9, 12\}$. For each econometric model, figures in bold specify the RMSFE-best model across all specifications, for a given forecast horizon. Grey boxes indicate the RMSFE-best models across all econometric models for a given forecast horizon.

Empirical findings unveil noteworthy insights. To start with, most entries in the first panel of Table 4.2 are below to one, indicating that the majority of model specifications outperform the naive benchmark during all forecast horizons. In addition, the majority of the competing models perform considerably better than the rolling mean benchmark, especially at horizons $h = \{1, 3\}$, with the bagging model exhibiting the best performance. The forecasting performance of the competing models tends to deteriorate relative to the benchmark model as the forecast horizon increases, while the bagging and the CSR models showing the best forecasting performance. Most importantly, the specifications that rely on either Google Trends data or Google Trends and News Trends data combined deliver the smallest relative forecast errors suggesting predictive superiority not only against the benchmark but also compared to models that utilise macroeconomic common factors. Findings also imply that the model specifications that employ Google Trends predictors provide significant gains especially for short-horizons and in some cases also for horizons $h = \{6, 9\}$ since the null hypothesis of equal forecast accuracy is rejected and generate the lowest ratios. This can be shown by the plethora of entries that are signed with asterisks, meaning Diebold-Mariano test rejections.

On a similar note, Google Trends predictors seem to provide the largest forecast benefits when forecasting the interquartile range of five-year price expectations as shown the second panel of Table 4.2. That is, the majority of the lowest forecast error ratios and bold entries are located within the Google Trends and News Trends section, with forecast improvements against the rolling mean benchmark being around 30-45% for one-step and three-step ahead forecasts, while for longer term forecasts accuracy decreases considerably. On the other hand, models relying only on macroeconomic indicators outperform the simple benchmark significantly only in one-step ahead horizons, with forecast benefits roughly up to 10%. Additionally, similarly to the findings of one-year price expectations, bagging-based specifications exhibit in overall the best forecasting performance.

Turning now to breakeven inflation rates, empirical results portray an interesting picture. As can be seen in the first panel of Table 4.3, most entries appear to be lower than one, indicating that the majority of the specifications outperform the benchmark when the target variable is the five-year breakeven inflation rate. Moreover, in a similar manner with the previous findings for consumers' price expectations, bagging models seems to provide the best forecast accuracy as reflected by the presence of the grey boxes within the bagging section. However, the model specifications that provide the highest forecast accuracy are not the Google-based models as in the case of consumer-based uncertainty shown in Table 4.2 but the specifications that rely solely on macroeconomic factors. Except for the one-step ahead forecasts, where the bagging-based model utilising Google Trends data exhibits the lowest ratio, the models employing macroeconomic predictors are the best performers. The forecast improvement of the macroeconomic-based bagging model relative to the benchmark is almost 30% for one-step-ahead predictions while for forecast horizons of $h = \{3, 6, 9, 12\}$ is around 19-22%. The significant outperformance of the macroeconomic predictors can also be observed by the amount of asterisks up to horizon $h = 6$ in the first panel of Table 4.3 suggesting a rejection of the null hypothesis in the Diebold and Mariano test.

In line with the findings presented for the five-year breakeven inflation rates, the second panel in Table 4.3 and the first panel in Table 4.4 confirm the predictive superiority of macroeconomic data when the target variable is the seven-year (BE7Y) and 10-year (BE10Y) breakeven inflation. It is only in the case of one-step-ahead forecasts that the methods solely utilising Google search data outperform the macroeconomic-based specifications. In this specific scenario, the forecast accuracy improvement against the benchmark is up to 30% with macroeconomic factors alone, while it climbs to around 55% for models based on Google search data. However, the forecasting performance of specifications based on Google and News Trends data series deteriorate rapidly as the forecast horizon increases with forecast improvements relative to the rolling mean benchmark declining to around 5-10%. On the other hand, the forecast improvement of the bagging-based models that use macroeconomic factors is around 19-23% and 14-18% for the seven-year and 10-year breakeven rates, respectively. It is worth mentioning that the statistical superiority of the macroeconomic-based models is stronger for the seven-year and 10-year breakeven inflation rates since we can find rejections of the null hypothesis of the Diebold and Mariano test up to horizons of $h = 12$ and $h = 9$, respectively.

Finally, the second panel of Table 4.4 offers noteworthy insights when focusing on the disagreement among professional forecasters. For shorter forecast horizons of up to $h = 3$, most competing models show only limited improvement over the naive benchmark model, as indicated by the majority of entries being close to one. Only the macroeconomic factors document in some cases significant forecast gains of around 8-12%. At longer forecast horizons of $h = \{9, 12\}$, Google search data demonstrate some gains in forecast accuracy, as evidenced by figures significantly lower than unity and occasional rejections of the null hypothesis in the Diebold and Mariano test. The bagging model consistently exhibits the highest forecast accuracy across these varying horizons, as reflected by the prevalence of grey boxes, with only one exception at $h = 1$ in which the random forest provides the lowest forecast error.

4.4.2 Time-varying Predictability

Interesting insights could also be revealed when we examine the predictability in the temporal dimension. Figures 4.3 - 4.8 plot cumulative squared forecast errors, normalised by the number of errors, for bagging-based models using different set of predictors and for selected forecast horizons. Specifically, Figure 4.3 shows the normalised cumulative forecast errors when the target variable is the interquartile range of consumers' one-year price expectations. The plots demonstrate that during the Global Financial Crisis of 2007-2009 the forecast errors of all model specifications increase significantly, highlighting the limits of forecasting models during turbulent times. Forecast errors decline rapidly as the U.S. economy recovers and shifts to an expansionary territory from 2010 up to 2019. Although the Covid-19 pandemic have not caused any significant uptick to forecast errors, the subsequent energy crisis and supply disruptions alongside with the effects of the unprecedented fiscal stimulus packages created a strong inflationary wave which led to an increased forecast errors over the period of 2021-2022. Also, it is evident that for horizons $h = \{3, 6\}$ the model which uses only a traditional set of macroeconomic predictors outperforms the rest of the models from the beginning of the out-of-sample exercise to either the early phase of the Covid-19 pandemic or the beginning of the energy crisis. This demonstrates that during a relatively stable inflationary period, a model that relies only on macroeconomic information tracks well developments related to inflation uncertainty. However, since the start of the pandemic, model specifications that utilise Google Trends data provide significant forecast benefits as the trend of the forecast error curves between Google-based and macroeconomic-based datasets changes in a material way leading the Google-based curves below the macroeconomic error curve. At the end of sample period, the model that uses Google search data, combined in some cases with News Trends data, exhibits the best forecasting performance against the other model specifications. It is worth noting that for the forecast horizon $h = 12$ the model that uses Google search outperforms after 2013 consistently over the out-of-sample period.

Figure 4.4 displays a pattern similar to that observed with one-year consumer-based price expectations, extending its applicability to five-year price expectations. A notable difference is that the model utilising Google and News Trends data begins to outperform the macroeconomic-based model around 2015-2017 for forecast horizons of $h = \{1, 3, 6\}$, while at the $h = 12$ horizon, this outperformance starts approximately in 2018-2019. Similar to the previous case, towards the end of the sample period, the forecast error curves of the Google-based models demonstrate greater accuracy compared to those of the macro-based models, particularly for forecast horizons of $h = 1$ and $h = 3$. Another distinction from the one-year price expectations case is that the forecast errors recorded during the 2008 crisis are comparable to those in the more recent inflationary period, except at the $h = 3$ horizon, where errors during the Global Financial Crisis were greater than those observed during 2021-2023.

Turning now to breakeven inflation rates, Figure 4.5 reveals some interesting insights.

For horizon $h = 1$, the Google-based bagging model outperforms the rest of the specifications from around 2012-2013 up to the end of the sample period when the target variable is the five-year breakeven inflation rate. However, for horizons $h = \{3, 6, 12\}$ the macroeconomic factors appear to provide the highest forecast gains on a consistent basis. Furthermore, we observe a substantial rise in forecast errors during the recession of 2008-2009 a trend that is in line with the case of consumers' uncertainty. However, in contrast with UMC1Y and UMC5Y, forecast errors do not appear to trend upwards during the recent spike in inflation uncertainty, showing that the relative performance of all model specifications remained stable under a volatile regime.

In line with the predictability of the five-year breakeven rate, Figures 4.6 - 4.7 exhibit a similar picture in which the majority of the forecast benefits for the macroeconomic-based bagging model for horizons higher than $h = 1$ occur shortly after the end of the Global Financial Crisis of 2007-2009, while at the one-step-ahead horizon the specification that uses only Google Trends set of predictors shows a forecast superiority from around 2012 onwards. Also, all model specifications document large forecast errors during the Global Financial Crisis of 2007-2009 while they remain relatively stable during the Covid-19 pandemic and the subsequent inflationary wave.

Finally, Figure 4.8 plots the normalised cumulative forecast errors for disagreement across professional forecasters. For prediction horizons of $h = 1$ and $h = 3$, the model based on macroeconomic factors seems to outperform by a small margin the models that utilise Google and News Trends predictors. The outperformance of macroeconomic data is slightly stronger for one-month ahead predictions. However, when we generate forecasts for $h = \{6, 12\}$ the Google search and News Trends datasets provide the best forecasting performance, especially for $h = \{12\}$ in which the gap between the forecast error curves widens significantly shortly after the beginning of our forecast experiments. Across all forecast horizons, forecast errors exhibit a significant spike over the 2009 recession but afterwards they are trending lower with a moderate increase during the Covid-19 pandemic and the subsequent energy crisis.

4.4.3 Measuring the Benefits of Google Trends

Considering that Google Trends series appear to be quite informative in forecasting consumer-based inflation uncertainty measures, a natural next step in our analysis is to explicitly measure their forecast gains by comparing the performance of bagging-based specifications that include Google Trends series against those that include only macroeconomic information or News Trends series. Table 4.5 shows the relative RMSFEs with values below unity suggesting a superiority of Google Trends series. It is clear that while forecasting measures of inflation uncertainty derived from consumer surveys (UMC1Y, UMC5Y), forecast benefits are observed across all horizons. However, these benefits progressively diminish as we move from short-term to long-term horizons. Additionally, it's noteworthy that the Google Trends data provide more significant gains in forecasting performance compared to macroeconomic variables, as opposed to the gains observed with News Trends data. Moving to the breakeven inflation rates, Google search series provide forecast gains, compared to macroeconomic predictors, only at one-step ahead horizons while for $h = \{3, 6, 9, 12\}$ they underperform in most cases by around 15%. Interestingly, their forecast power remain again relatively stable when compared to specifications based on News Trends data. Finally, in the case of professional forecasters' uncertainty, Google search terms appear to offer considerable numerical benefits compared to macroeconomic predictors only at $h = 12$ with gains reaching almost up to 25%. However, when compared to News Count predictors, Google search data appear to add considerable forecast benefits at $h = \{6, 9, 12\}$ horizons, with gains ranging from around 10% to roughly 30%.

4.4.4 Which Google Search Series Have Predictive Power?

Taking into consideration that we have already established the usefulness of Google Trends series in predicting indicators of inflation uncertainty, in the final section of our empirical analysis we attempt to identify from which predictors the forecasting power is coming from. To facilitate this, the out-of-sample R^2 is employed, as introduced by Campbell and Thompson (2008) and defined in Equation (4.7), to evaluate the predictive power from the Google search series. Using a rolling window of 48 observations and the same out-of-sample period as in the main forecasting exercise, each measure of inflation uncertainty is regressed upon every Google search series separately, that is $Y_{j,t} = \alpha_i + \beta_i X_{i,t-h} + \epsilon_t$ where j denotes the different measures of inflation uncertainty and i reflects the individual Google Trends series. Table 4.6 shows the R^2_{OoS} across the target variables and forecast horizons. At every panel, the search queries are ranked according to the average R^2_{OoS}. The entries in this table confirm the predictive power of Google Trends when forecasting a consumer-based inflation uncertainty measure as reflected by the large R^2_{OoS} values in the first two panels compared to the rest of the table. In particular, the R^2_{OoS} range from 0.34 to 0.64 across all terms and forecast horizons while the average does not fall below 0.48 for any query, highlighting the significant predictive power of Google Trends for UMC1Y and UMC5Y. Moving to the breakeven inflation rate panels, it is clear that the R^2_{OoS} decreases ranging from 0.14 to 0.55 while the average figures show R^2_{OoS} values as low as 0.25. Finally, the last panel of Table 4.6 reflects the limited forecast benefits that emerge when utilising Google search series to forecast a measure of inflation uncertainty based on surveys from professional forecasters.

Turning now to the selection of the Google search terms, when forecasting one- and five-year consumer-based price expectations (UMC1Y, UMC5Y), it is unsurprising to find a notable overlap in the search queries with the highest R^2_{OoS}, predominantly centered around queries related to "inflation". On a similar note, when targeting a market-based inflation uncertainty indicator (BE5Y, BE7Y, BE10Y) there is a significant commonality between the selected keywords. However, in this case, the dominance of inflation-related keywords is tempered by the presence of additional terms such as "interest rate" and "economy".

Finally, it is worth commenting on the evolution of the R^2_{OoS} entries over the forecast horizons. The first two panels that target the consumer-based uncertainty measures show a declining trend in the R^2_{OoS} as the forecast horizon increases. This is in line with the findings shown in Table 4.2 and reflect a diminishing predictive power of Google search series when producing long horizon predictions. However, interesting insights are revealed when forecasting breakeven inflation rates. Table 4.6 shows that the individual predictive power of the top five queries increases since the R^2_{OoS} values are trending upwards as the prediction horizon increases, a conclusion which does not corroborate the findings presented in Table 4.3 and show that when all Google indicators are included within the bagging, complete subset regressions and random forest specifications the forecasting power of Google Trends deteriorates or remains relatively stable as the prediction horizon increases.

4.5 Concluding Remarks

In conclusion, this study aims to explore the efficacy of alternative predictors derived from internet search and news articles in forecasting inflation uncertainty in the United States. A set of predictors using Google Trends and Bloomberg's News Trends is constructed, focusing on inflation-related keywords. Considering that an objective measure of inflation uncertainty does not exist, three distinct measures of uncertainty are computed to capture the variability in price expectations among households, investors, and professional forecasters.

Our empirical findings reveal several interesting insights. First, when forecasting households' inflation uncertainty, Google Trends predictors exhibit remarkable predictability, and outperform significantly the macroeconomic predictors particularly at horizons ranging from one to three months ahead. However, the predictive accuracy of Google Trends predictors diminishes as the forecasting horizon extends, eventually converging with the performance of traditional factors. The combination of Google and News Trends also yields strong forecast accuracy in certain cases. In contrast, when forecasting investors' inflation uncertainty, macroeconomic factors consistently provide the most accurate predictions across all horizons except for the one-month forecast. Second, our analysis reveals that the bagging specifications in most cases deliver the most accurate predictions, with the exception of consumer surveys, where both bagging and complete subset regression models perform the best compared to the benchmark.

Third, our investigation into predictability over time highlights intriguing patterns. For forecasting households' uncertainty, macroeconomic factors appear to be the most informative data source until the period of around 2017-2019. However, during the Covid-19 pandemic and times of heightened uncertainty due to geopolitical tensions and supply shocks, Google and News Trends exhibit significant forecasting gains. On the other hand, when predicting investors' uncertainty, macroeconomic factors consistently outperform alternative predictors over the out-of-sample period for horizons ranging from three months to one year ahead.

Lastly, we seek to shed light on which Google Trends series contain the highest predictive power. When the target variable is based on consumer surveys, Google Trends queries predominantly related to "inflation" are selected as the most informative. However, when uncertainty measures are based on market indicators and professional forecasters, the selected queries include additional terms related to the "economy" and "interest rates."

In summary, our research suggests that Google and News Trends data could capture relatively well the behavior of consumers and corresponding survey-based uncertainty indicators. However, when it comes to predicting market-based measures of uncertainty, their contribution to forecasting gains is limited. These findings align with the conclusions and insights presented in Carroll (2003), which suggest that not everyone actively follows inflation-related articles. Nevertheless, when individuals specifically search for inflation information on Google, it signifies their interest and enhances their likelihood of utilising the acquired information to update their expectations and confidence levels regarding inflation.

Table 4.2: Forecasting One- and Five-Year Univ. of Michigan Survey Price Expectations

Model	Predictors	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
		Panel A: UMC1Y				
Bagging	MF	0.869**	0.901	0.957	0.958	0.970
	NT	0.710***	0.875*	0.933	1.051	1.042
	GT	**0.591*****	**0.690****	0.806**	**0.832***	**0.898**
	GTNT	0.596***	0.700**	0.792*	0.858*	0.959
CSR	MF	0.906*	0.928	0.958	0.956	0.976
	NT	0.887**	0.922	0.980	1.014	1.013
	GT	0.663***	0.697**	0.784*	0.880	0.941
	GTNT	0.655***	0.702**	**0.768***	0.886	0.966
RFs	MF	0.872**	0.902	0.922	0.939	0.980
	NT	0.837**	0.898	0.978	0.992	0.992
	GT	0.627***	0.706**	0.844*	0.884	0.920
	GTNT	0.636***	0.721**	0.832*	0.899	0.943
		Panel B: UMC5Y				
Bagging	MF	0.896*	0.927	0.972	0.994	0.968
	NT	0.702***	0.837*	0.954	0.998	0.968
	GT	0.559***	0.684**	0.810*	0.823	0.949
	GTNT	**0.556*****	**0.674****	**0.807***	**0.812***	**0.935**
CSR	MF	0.931	0.950	0.982	0.984	0.978
	NT	0.863*	0.901	0.976	0.987	0.975
	GT	**0.626*****	**0.677*****	**0.788***	0.881	0.965
	GTNT	0.640***	0.680***	0.792*	**0.858**	**0.942**
RFs	MF	0.885**	0.899	0.914	0.928	0.963
	NT	0.795**	0.855*	0.954	0.945	0.948
	GT	**0.591*****	**0.719****	**0.841***	**0.867**	**0.927**
	GTNT	0.607***	0.736**	0.878	0.883	0.930

Notes: Entries in the table show relative Root Mean Squared Forecast Errors (RMSFEs) of each model relative to the benchmark model for forecasting the interquartile range of the one-year (Panel A) and five-year (Panel B) price expectations based on University of Michigan Survey of Consumer Expectations. Therefore, entries lower than unity imply that the model of reference that employs a particular set of predictors outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon and econometric model. Grey boxes reflect the lowest relative error achieved considering all competing models for a specific forecast horizon. The models are estimated by employing a rolling window scheme that considers a sample period from January 2004 to December 2022, consisting of 48 observations. Predictions are generated for the following monthly horizons, $h = \{1, 3, 6, 9, 12\}$. Values followed by asterisks (***1%, **5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table 4.3: Forecasting Five- and Seven-Year Breakeven Inflation Rates

Model	Predictors	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
		Panel A: BE5Y				
Bagging	MF	0.715***	**0.782****	**0.819***	**0.811**	**0.813**
	NT	0.573***	0.998	1.042	1.058	1.057
	GT	**0.505*****	0.893	0.936	0.920	0.925
	GTNT	0.524***	0.927	0.974	0.959	0.995
CSR	MF	**0.769*****	**0.813****	**0.857***	**0.857**	**0.887**
	NT	0.971*	0.985	1.004	1.010	1.010
	GT	0.826***	0.894	0.943	0.938	0.900
	GTNT	0.836***	0.903	0.964	0.968	0.938
RFs	MF	**0.782*****	**0.835****	**0.903**	0.965	1.025
	NT	0.954*	1.002	1.046	1.067	1.070
	GT	0.785***	0.899	0.967	**0.964**	**0.911**
	GTNT	0.796***	0.929	1.011	1.007	0.949
		Panel B: BE7Y				
Bagging	MF	0.707***	**0.777****	**0.813***	**0.794***	**0.787***
	NT	0.619***	0.989	1.034	1.045	1.043
	GT	**0.548*****	0.890	0.940	0.937	0.953
	GTNT	0.566***	0.929	0.974	0.972	0.981
CSR	MF	**0.769*****	**0.812****	**0.852***	**0.848**	**0.874**
	NT	0.969**	0.986	1.002	1.004	1.003
	GT	0.823***	0.893	0.941	0.953	0.913
	GTNT	0.834***	0.903	0.960	0.978	0.955
RFs	MF	**0.773*****	**0.829****	**0.898**	**0.962**	1.019
	NT	0.943**	1.001	1.037	1.059	1.061
	GT	0.783***	0.892	0.968	0.979	**0.933**
	GTNT	0.803***	0.923	1.010	1.017	0.977

Notes: Entries in the table show relative Root Mean Squared Forecast Errors (RMSFEs) of each model relative to the benchmark model for forecasting the five-year (Panel A) and seven-year (Panel B) breakeven inflation rates. Therefore, entries lower than unity imply that the model of reference that employs a particular set of predictors outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon and econometric model. Grey boxes reflect the lowest relative error achieved considering all competing models for a specific forecast horizon. The models are estimated by employing a rolling window scheme that considers a sample period from January 2004 to December 2022, consisting of 48 observations. Predictions are generated for the following monthly horizons, $h = \{1, 3, 6, 9, 12\}$. Values followed by asterisks (***1%, **5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

Table 4.4: Forecasting 10-Year Breakeven Infl. and Disagreement of Prof. Forecasters

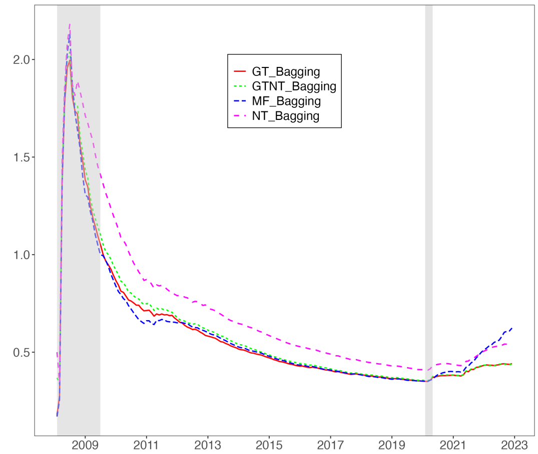
Model	Predictors	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
		Panel A: BE10Y				
Bagging	MF	0.740***	**0.818***	**0.857***	**0.848***	**0.818**
	NT	0.575***	0.997	1.034	1.052	1.043
	GT	**0.540*****	0.912	0.955	0.936	0.977
	GTNT	0.548***	0.942	0.982	0.970	1.001
CSR	MF	**0.796*****	**0.840****	**0.878***	**0.867**	**0.870**
	NT	0.969**	0.991	1.005	1.010	1.007
	GT	0.819***	0.894	0.947	0.945	0.946
	GTNT	0.827***	0.900	0.957	0.974	0.976
RFs	MF	0.812***	**0.854****	**0.898**	**0.939**	**0.963**
	NT	0.950*	1.010	1.034	1.050	1.050
	GT	**0.777*****	0.885	0.978	0.983	0.974
	GTNT	0.789***	0.920	1.000	1.013	0.987
		Panel B: SPF				
Bagging	MF	**0.933***	**0.880***	**0.897***	0.899	0.992
	NT	0.991	1.023	1.024	1.009	1.060
	GT	0.949	0.996	0.907	**0.836***	**0.759***
	GTNT	0.966	0.996	0.909	0.861	0.829
CSR	MF	**0.932***	**0.925**	**0.922**	0.951	1.029
	NT	0.987	0.993	1.002	0.988	1.018
	GT	0.943	0.971	0.935	**0.882**	**0.823**
	GTNT	0.956	0.971	0.942	0.900	0.859
RFs	MF	**0.929***	**0.982**	1.047	1.084	1.154
	NT	0.993	1.015	1.034	1.018	1.043
	GT	0.954	0.989	**0.927**	**0.844***	**0.777***
	GTNT	0.948	0.988	0.939	0.868	0.817

Notes: Entries in the table show relative Root Mean Squared Forecast Errors (RMSFEs) of each model relative to the benchmark model for forecasting the 10-year breakeven inflation rates (Panel A) and the disagreement among professional forecasters (Panel B). Therefore, entries lower than unity imply that the model of reference that employs a particular set of predictors outperforms the benchmark. Numbers in bold indicate the most accurate model for each forecast horizon and econometric model. Grey boxes reflect the lowest relative error achieved considering all competing models for a specific forecast horizon. The models are estimated by employing a rolling window scheme that considers a sample period from January 2004 to December 2022, consisting of 48 observations. Predictions are generated for the following monthly horizons, $h = \{1, 3, 6, 9, 12\}$. Values followed by asterisks (***1%, **5% level, *10% level) are significantly superior to the benchmark model based on the Diebold-Mariano test.

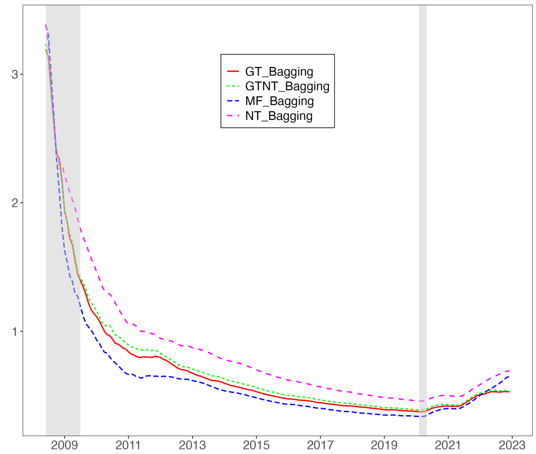
Table 4.5: Google Trends Against Macroeconomic and News Trends Predictors

Target	Predictors	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
UMC1Y	MF	0.680	0.766	0.842	0.869	0.925
	NT	0.833	0.789	0.864	0.792	0.862
UMC5Y	MF	0.624	0.738	0.833	0.828	0.981
	NT	0.796	0.817	0.849	0.825	0.981
BE5Y	MF	0.706	1.142	1.143	1.135	1.137
	NT	0.881	0.895	0.899	0.869	0.875
BE7Y	MF	0.775	1.145	1.156	1.181	1.211
	NT	0.885	0.900	0.909	0.896	0.914
BE10Y	MF	0.729	1.114	1.115	1.103	1.194
	NT	0.939	0.914	0.924	0.889	0.937
SPF	MF	1.017	1.132	1.011	0.929	0.765
	NT	0.957	0.973	0.886	0.829	0.716

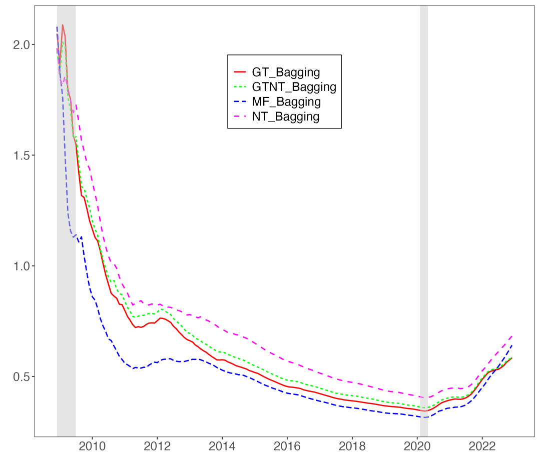
Notes: Entries in the table show relative Root Mean Squared Forecast Errors (RMSFEs) of Google-based bagging models relative to bagging models that use as predictors the macroeconomic factors (MF) data News Trends (NT) data. Therefore, entries lower than unity imply that the Google-based model outperforms the bagging models that employ macroeconomic factors and News Trends predictors. The models are estimated by employing a rolling window scheme that considers a sample period from January 2004 to December 2022, consisting of 48 observations. Predictions are generated for the following monthly horizons, $h = \{1, 3, 6, 9, 12\}$.



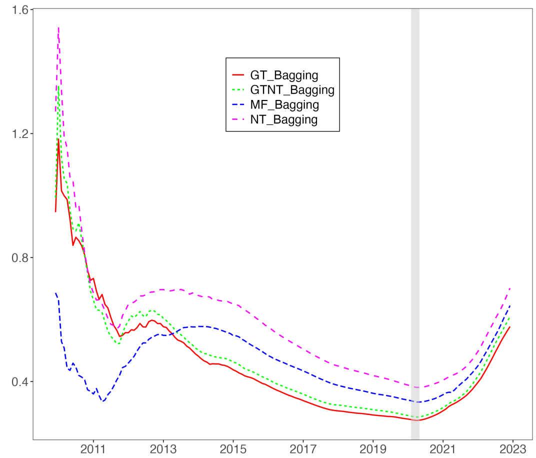
(a) Forecast Horizon: $h = 1$



(b) Forecast Horizon: $h = 3$



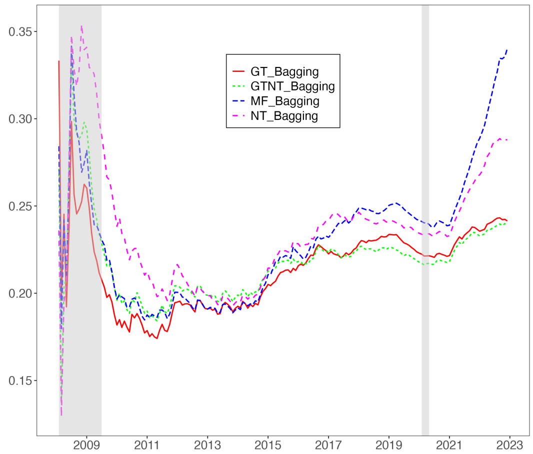
(c) Forecast Horizon: $h = 6$



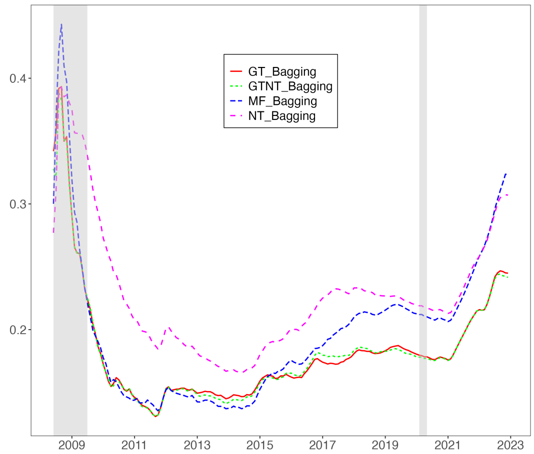
(d) Forecast Horizon: $h = 12$

Figure 4.3: Normalised Cumulative Forecast Errors for UMC1Y

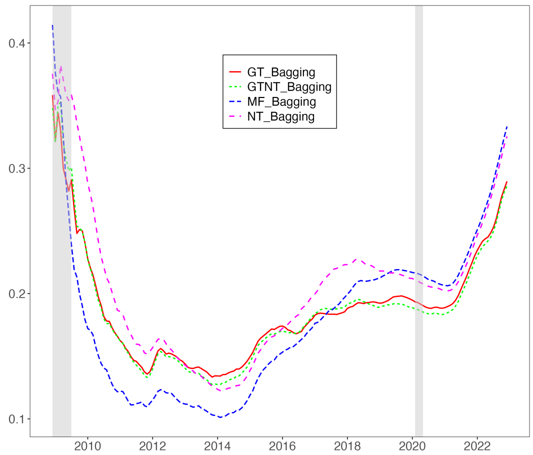
Notes: This graph shows the cumulative root squared forecast error of various bagging-based specifications across different set of predictors and forecast horizons, normalised by the number of errors. The target variable is the UMC1Y which denotes the interquartile range of the one-year price expectations of University of Michigan Survey of Consumer Expectations. The red solid line shows the cumulative forecast error of the bagging model based on Google Trends (GT) data, the green dotted line exhibits the forecast errors based on Google Trends (GT) and News Trends (NT) data, the dashed blue line denotes the forecast errors based on macroeconomic factors (MF), while the pink long-dashed line reflects the forecast errors of a model based only on News Trends predictors. Shaded areas reflect NBER-dated recessions of the U.S. economy.



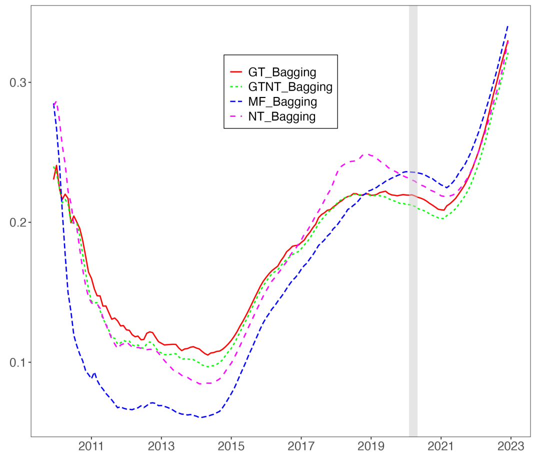
(a) Forecast Horizon: $h = 1$



(b) Forecast Horizon: $h = 3$



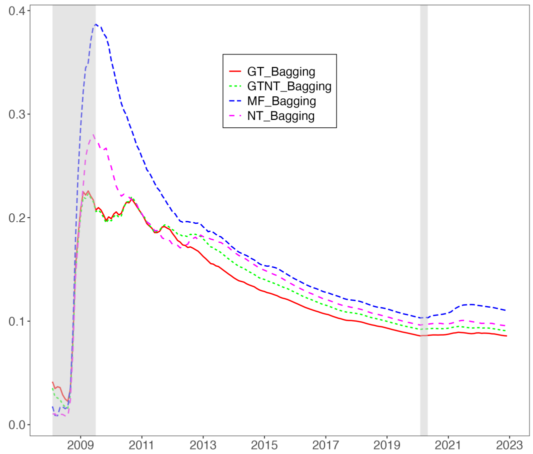
(c) Forecast Horizon: $h = 6$



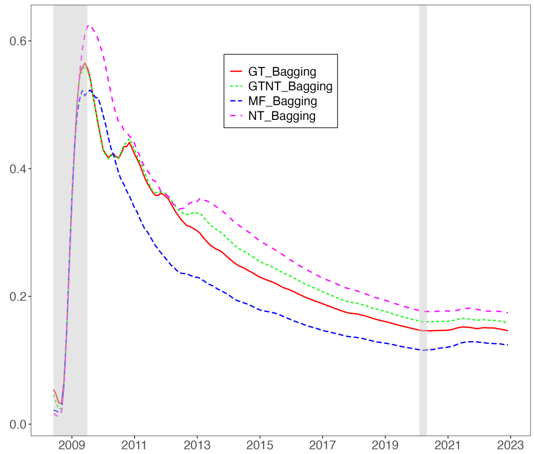
(d) Forecast Horizon: $h = 12$

Figure 4.4: Normalised Cumulative Forecast Errors for UMC5Y

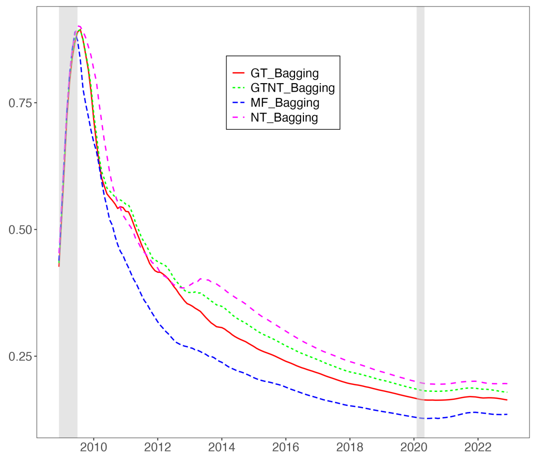
Notes: This graph shows the cumulative root squared forecast error of various bagging-based specifications across different set of predictors and forecast horizons, normalised by the number of errors. The target variable is the UMC5Y which denotes the interquartile range of the five-year price expectations of University of Michigan Survey of Consumer Expectations. The red solid line shows the cumulative forecast error of the bagging model based on Google Trends (GT) data, the green dotted line exhibits the forecast errors based on Google Trends (GT) and News Trends (NT) data, the dashed blue line denotes the forecast errors based on macroeconomic factors (MF), while the pink long-dashed line reflects the forecast errors of a model based only on News Trends predictors. Shaded areas reflect NBER-dated recessions of the U.S. economy.



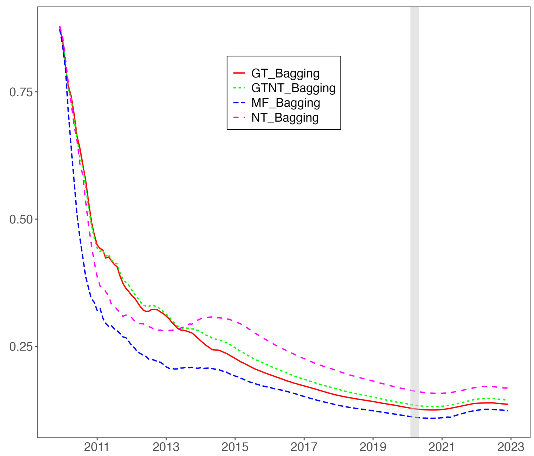
(a) Forecast Horizon: $h = 1$



(b) Forecast Horizon: $h = 3$



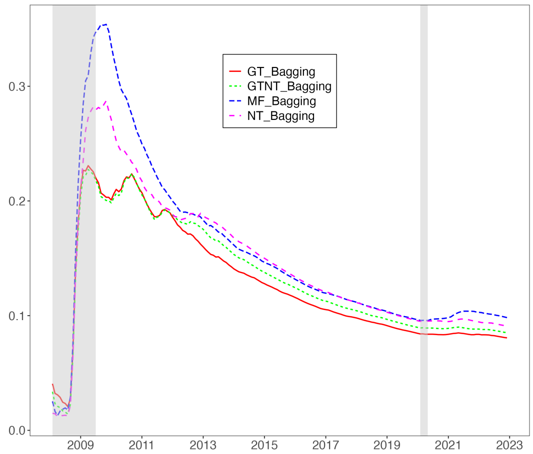
(c) Forecast Horizon: $h = 6$



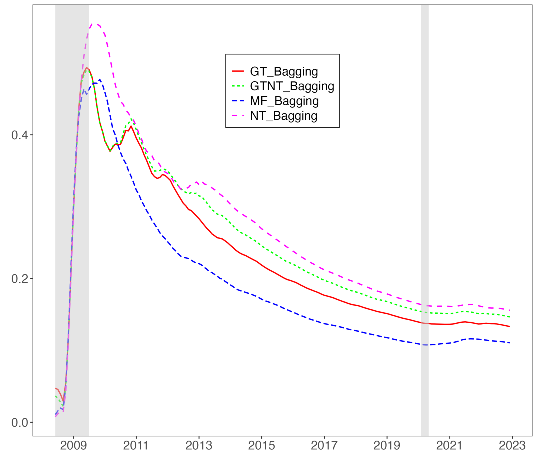
(d) Forecast Horizon: $h = 12$

Figure 4.5: Normalised Cumulative Forecast Errors for BE5Y

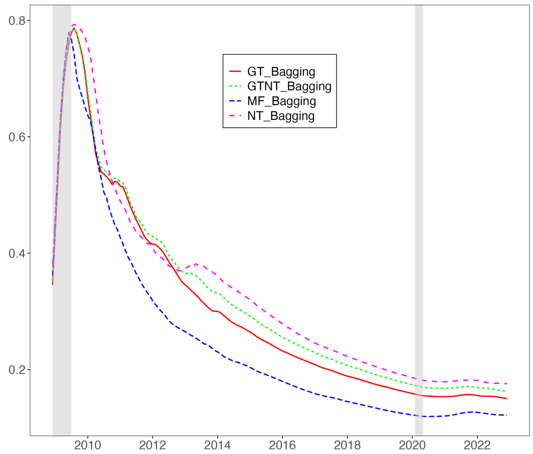
Notes: This graph shows the cumulative root squared forecast error of various bagging-based specifications across different set of predictors and forecast horizons, normalised by the number of errors. The target variable is the BE5Y which denotes the five-year breakeven inflation rates. The red solid line shows the cumulative forecast error of the bagging model based on Google Trends (GT) data, the green dotted line exhibits the forecast errors based on Google Trends (GT) and News Trends (NT) data, the dashed blue line denotes the forecast errors based on macroeconomic factors (MF), while the pink long-dashed line reflects the forecast errors of a model based only on News Trends predictors. Shaded areas reflect NBER-dated recessions of the U.S. economy.



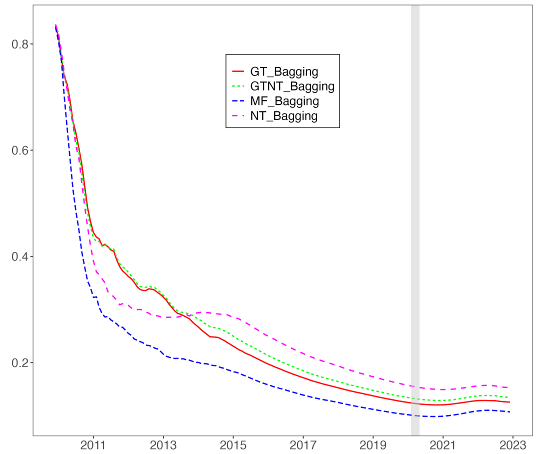
(a) Forecast Horizon: $h = 1$



(b) Forecast Horizon: $h = 3$



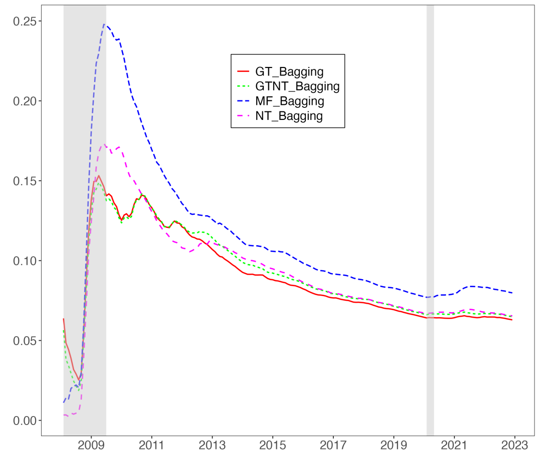
(c) Forecast Horizon: $h = 6$



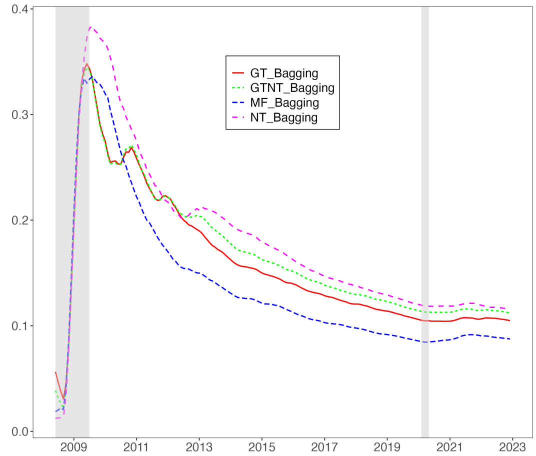
(d) Forecast Horizon: $h = 12$

Figure 4.6: Normalised Cumulative Forecast Errors for BE7Y

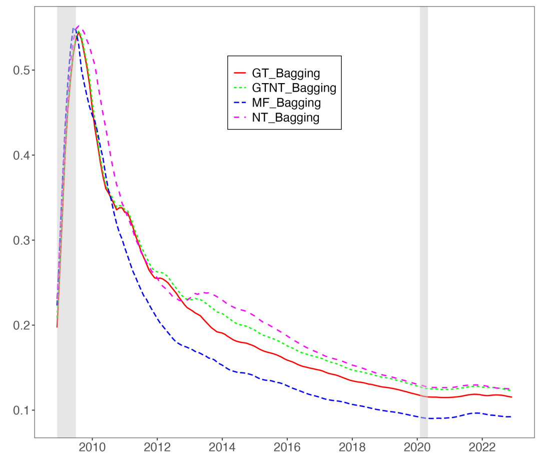
Notes: This graph shows the cumulative root squared forecast error of various bagging-based specifications across different set of predictors and forecast horizons, normalised by the number of errors. The target variable is the BE7Y which denotes the seven-year breakeven inflation rates. The red solid line shows the cumulative forecast error of the bagging model based on Google Trends (GT) data, the green dotted line exhibits the forecast errors based on Google Trends (GT) and News Trends (NT) data, the dashed blue line denotes the forecast errors based on macroeconomic factors (MF), while the pink long-dashed line reflects the forecast errors of a model based only on News Trends predictors. Shaded areas reflect NBER-dated recessions of the U.S. economy.



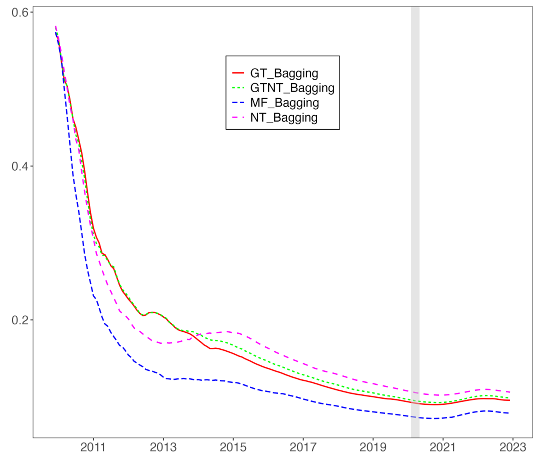
(a) Forecast Horizon: $h = 1$



(b) Forecast Horizon: $h = 3$



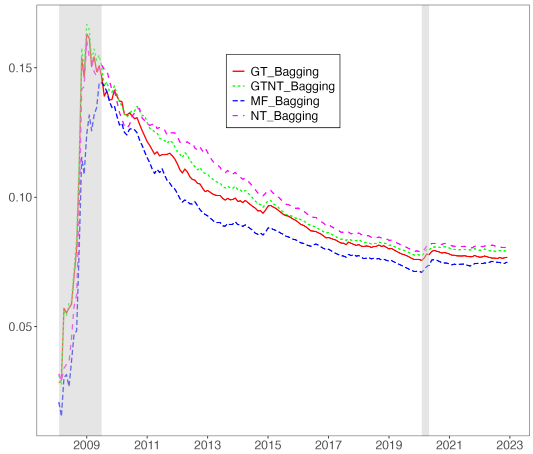
(c) Forecast Horizon: $h = 6$



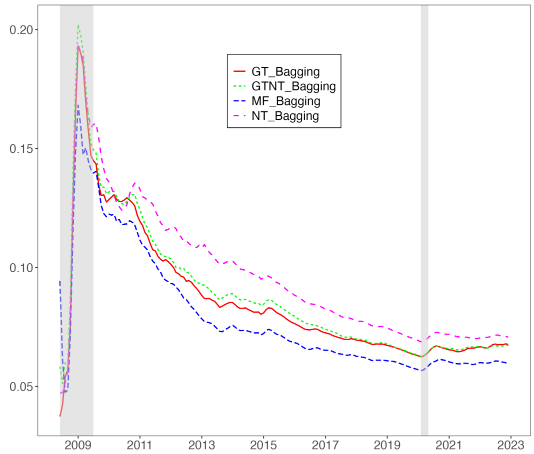
(d) Forecast Horizon: $h = 12$

Figure 4.7: Normalised Cumulative Forecast Errors for BE10Y

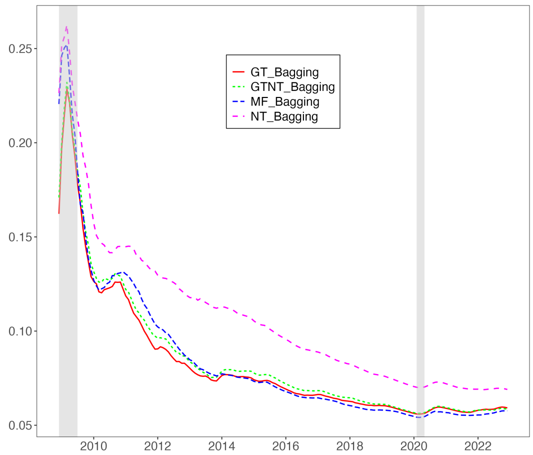
Notes: This graph shows the cumulative root squared forecast error of various bagging-based specifications across different set of predictors and forecast horizons, normalised by the number of errors. The target variable is the BE10Y which denotes the 10-year breakeven inflation rates. The red solid line shows the cumulative forecast error of the bagging model based on Google Trends (GT) data, the green dotted line exhibits the forecast errors based on Google Trends (GT) and News Trends (NT) data, the dashed blue line denotes the forecast errors based on macroeconomic factors (MF), while the pink long-dashed line reflects the forecast errors of a model based only on News Trends predictors. Shaded areas reflect NBER-dated recessions of the U.S. economy.



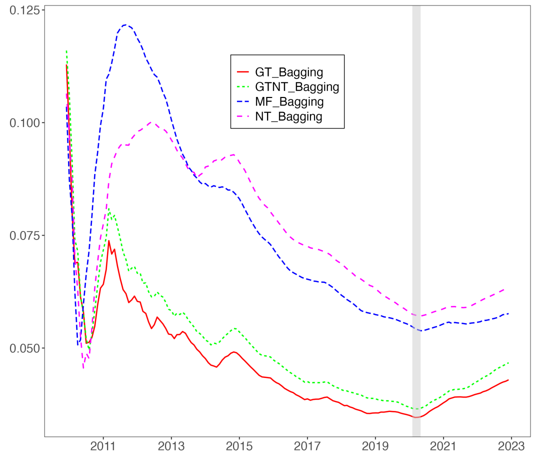
(a) Forecast Horizon: $h = 1$



(b) Forecast Horizon: $h = 3$



(c) Forecast Horizon: $h = 6$



(d) Forecast Horizon: $h = 12$

Figure 4.8: Normalised Cumulative Forecast Errors for SPF

Notes: This graph shows the cumulative root squared forecast error of various bagging-based specifications across different set of predictors and forecast horizons, normalised by the number of errors. The target variable is the SPF which denotes the disagreement among professional forecasters computed as the interquartile range from monthly point predictions. The red solid line shows the cumulative forecast error of the bagging model based on Google Trends (GT) data, the green dotted line exhibits the forecast errors based on Google Trends (GT) and News Trends (NT) data, the dashed blue line denotes the forecast errors based on macroeconomic factors (MF), while the pink long-dashed line reflects the forecast errors of a model based only on News Trends predictors. Shaded areas reflect NBER-dated recessions of the U.S. economy.

Table 4.6: Google Trends Individual Predictive Power According to R^2_{OoS}

	$h=1$	$h=3$	$h=6$	$h=9$	$h=12$	Average
			Panel A: UMC1Y			
"what is inflation"	0.53	0.55	0.57	0.50	0.44	0.52
"inflation rates"	0.55	0.56	0.57	0.47	0.37	0.51
"us inflation"	0.52	0.56	0.58	0.48	0.38	0.50
"dollar inflation"	0.48	0.52	0.54	0.48	0.45	0.50
"current inflation"	0.52	0.54	0.53	0.43	0.36	0.48
			Panel B: UMC5Y			
"us inflation"	0.63	0.64	0.55	0.47	0.40	0.54
"us inflation rate"	0.59	0.60	0.52	0.42	0.34	0.49
"what is inflation"	0.56	0.54	0.49	0.42	0.40	0.48
"current inflation"	0.57	0.57	0.49	0.43	0.35	0.48
"dollar inflation"	0.55	0.53	0.46	0.41	0.42	0.48
			Panel C: BE5Y			
"interest rate"	0.39	0.46	0.44	0.41	0.48	0.44
"economy"	0.38	0.43	0.44	0.40	0.42	0.41
"inflation"	0.24	0.32	0.41	0.46	0.55	0.40
"what is inflation"	0.19	0.25	0.29	0.30	0.37	0.28
"interest rates"	0.14	0.20	0.25	0.28	0.40	0.25
			Panel D: BE7Y			
"interest rate"	0.40	0.47	0.46	0.44	0.48	0.45
"economy"	0.36	0.42	0.45	0.44	0.45	0.42
"inflation"	0.23	0.31	0.40	0.45	0.54	0.39
"what is inflation"	0.19	0.25	0.29	0.31	0.37	0.28
"inflation rates"	0.22	0.26	0.24	0.25	0.27	0.25
			Panel E: BE10Y			
"interest rate"	0.33	0.41	0.40	0.37	0.46	0.39
"economy"	0.32	0.38	0.40	0.39	0.41	0.38
"inflation"	0.18	0.26	0.36	0.41	0.51	0.34
"what is inflation"	0.19	0.26	0.30	0.30	0.36	0.28
"inflation calculator"	0.24	0.25	0.28	0.26	0.27	0.26
			Panel F: SPF			
"interest rates"	0.12	0.13	0.15	0.23	0.23	0.17
"us inflation"	0.08	0.18	0.17	0.20	0.21	0.17
"inflation rate"	0.10	0.14	0.16	0.15	0.21	0.15
"economy"	0.16	0.19	0.21	0.12	0.07	0.15
"inflation"	0.10	0.18	0.15	0.12	0.14	0.14

Notes: Entries in this table show the out-of-sample R^2 as defined in Campbell and Thompson (2008). The models under consideration here are linear univariate specifications of the form $Y_{j,t} = \alpha_i + \beta_i X_{i,t-h} + \epsilon_t$ where X_i reflects each individual Google search predictor. Each panel corresponds to a different inflation uncertainty measure and presents the top five search queries ranked according to the average R^2_{OoS} across forecasting horizons.

C Appendix

Table C.1: Complete Subset Regressions with $q = 4$

	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
			Panel A: UMC1Y		
MF	0.880	0.911	0.960	0.954	0.986
NT	0.848	0.889	0.965	1.020	1.021
GT	0.630	0.668	0.754	0.852	0.929
GTNT	0.631	0.682	0.739	0.860	0.965
			Panel B: UMC5Y		
MF	0.891	0.911	0.959	0.967	0.968
NT	0.812	0.857	0.957	0.980	0.966
GT	0.573	0.632	0.750	0.849	0.957
GTNT	0.590	0.641	0.756	0.823	0.932
			Panel C: BE5Y		
MF	0.683	0.742	0.797	0.799	0.860
NT	0.961	0.983	1.010	1.016	1.015
GT	0.785	0.879	0.940	0.928	0.886
GTNT	0.802	0.895	0.969	0.968	0.939
			Panel D: BE7Y		
MF	0.682	0.739	0.786	0.779	0.832
NT	0.959	0.984	1.006	1.007	1.006
GT	0.782	0.876	0.937	0.948	0.905
GTNT	0.800	0.897	0.961	0.983	0.961
			Panel E: BE10Y		
MF	0.720	0.781	0.827	0.811	0.827
NT	0.961	0.991	1.010	1.016	1.013
GT	0.777	0.878	0.944	0.936	0.939
GTNT	0.794	0.895	0.956	0.975	0.981
			Panel F: SPF		
MF	0.926	0.920	0.915	0.958	1.073
NT	0.987	0.993	1.003	0.982	1.023
GT	0.939	0.985	0.929	0.872	0.805
GTNT	0.959	0.982	0.940	0.890	0.847

Notes: Entries in this table show relative Root Mean Squared Forecast Errors (RMSFEs) of CSR model specifications relative to the benchmark model across several forecasting horizons. Each panel corresponds to a different inflation uncertainty measure.

Table C.2: Complete Subset Regressions with $q = 5$

	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
			Panel A: UMC1Y		
MF	0.858	0.903	0.968	0.965	1.016
NT	0.819	0.862	0.950	1.026	1.029
GT	0.606	0.652	0.741	0.834	0.921
GTNT	0.614	0.672	0.725	0.842	0.965
			Panel B: UMC5Y		
MF	0.848	0.868	0.921	0.947	0.965
NT	0.773	0.821	0.938	0.975	0.959
GT	0.546	0.608	0.730	0.827	0.951
GTNT	0.566	0.619	0.735	0.799	0.926
			Panel C: BE5Y		
MF	0.616	0.685	0.749	0.761	0.873
NT	0.955	0.983	1.016	1.023	1.020
GT	0.761	0.872	0.942	0.920	0.881
GTNT	0.782	0.898	0.978	0.968	0.944
			Panel D: BE7Y		
MF	0.613	0.681	0.731	0.726	0.823
NT	0.952	0.985	1.011	1.011	1.009
GT	0.757	0.869	0.938	0.943	0.904
GTNT	0.780	0.902	0.967	0.988	0.970
			Panel E: BE10Y		
MF	0.661	0.736	0.786	0.768	0.812
NT	0.955	0.995	1.018	1.024	1.019
GT	0.752	0.873	0.947	0.929	0.935
GTNT	0.778	0.901	0.961	0.976	0.986
			Panel F: SPF		
MF	0.932	0.931	0.931	0.995	1.144
NT	0.990	0.995	1.004	0.977	1.027
GT	0.944	1.003	0.930	0.874	0.802
GTNT	0.971	1.002	0.945	0.889	0.848

Notes: Entries in this table show relative Root Mean Squared Forecast Errors (RMSFEs) of CSR model specifications relative to the benchmark model across several forecasting horizons. Each panel corresponds to a different inflation uncertainty measure.

Table C.3: Bagging with $t_{crit} = 1.96$

	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Panel A: UMC1Y					
MF	0.853	0.893	0.962	0.969	0.986
NT	0.692	0.859	0.923	1.051	1.040
GT	0.579	0.682	0.805	0.815	0.874
GTNT	0.582	0.700	0.796	0.839	0.944
Panel B: UMC5Y					
MF	0.881	0.909	0.968	1.001	0.975
NT	0.682	0.818	0.948	0.997	0.965
GT	0.550	0.674	0.807	0.817	0.940
GTNT	0.552	0.668	0.804	0.795	0.922
Panel C: BE5Y					
MF	0.682	0.754	0.803	0.788	0.805
NT	0.567	0.990	1.042	1.056	1.054
GT	0.492	0.881	0.926	0.897	0.911
GTNT	0.518	0.920	0.967	0.944	0.951
Panel D: BE7Y					
MF	0.671	0.748	0.797	0.768	0.766
NT	0.566	0.984	1.035	1.044	1.039
GT	0.504	0.879	0.929	0.919	0.940
GTNT	0.525	0.927	0.966	0.963	0.974
Panel E: BE10Y					
MF	0.713	0.795	0.842	0.825	0.791
NT	0.573	0.992	1.037	1.053	1.043
GT	0.526	0.895	0.941	0.918	0.966
GTNT	0.541	0.933	0.972	0.960	0.996
Panel F: SPF					
MF	0.911	0.871	0.894	0.917	1.079
NT	0.993	1.023	1.022	0.994	1.048
GT	0.944	1.015	0.911	0.839	0.772
GTNT	0.972	1.022	0.917	0.868	0.844

Notes: Entries in this table show relative Root Mean Squared Forecast Errors (RMSFEs) of bagging model specifications relative to the benchmark model across several forecasting horizons. Each panel corresponds to a different inflation uncertainty measure.

Chapter 5

Conclusion

5.1 Concluding Remarks

This thesis examines the usefulness of alternative data in forecasting key macroeconomic indicators. In Chapter 2 we study the predictive power of Google search data in nowcasting GDP growth rates in the United States and Brazil. We find that the highest accuracy is achieved when Google Trends are combined with traditional economic predictors, suggesting that Google data appear to complement rather than to replace the macroeconomic variables provided by government agencies. Additionally, our findings suggest that the benefits of using search data are broadly similar for the United States and Brazil and depend on the variable-selection method applied to the factor model. The variable-selection method that yields the lowest forecast error varies not only across countries but also across forecast horizons, implying that preemptively choosing a specific method for variable selection proves to be quite challenging. The benefits of implementing variable selection are greater during one-step-ahead forecasts while their performance declines as we move to nowcast and backcast predictions.

Next, in Chapter 3, we explore the usefulness of internet search data in forecasting private consumption and its components, in the United States. We find that Google data consists of a valuable source when forecasting the aggregate private consumption expenditures, especially over longer-term horizons. Substantial gains are also observed when forecasting the consumption components separately. Google-based models are notably effective in predicting durable goods expenditures, particularly in the post-pandemic era, while for nondurable goods and services, models that utilise search data demonstrate some predictive power over the medium-term horizons. Moreover, we provide evidence that models that impose sparsity consistently outperform those using dense modelling, with random forests showing, in overall, the best forecasting performance, followed closely by LASSO specifications. Furthermore, an analysis of model performance over time indicates a general decline in accuracy during the early stages of the pandemic, a pattern typically observed in econometric models during periods of crisis and recovery.

Finally, in Chapter 4 we investigate the role of alternative data sources such as Google Trends and News Trends in forecasting various measures of inflation uncertainty. Our empirical results suggest that Google Trends predictors are highly effective for short-term forecasts (up to three months ahead) of households' inflation uncertainty, with bagging and complete subset regressions being the most successful methods. However, their forecasting accuracy diminishes over longer horizons, aligning with macroeconomic factors. News Trends also show satisfactory performance, and in some cases they provide higher forecast accuracy when combined with Google Trends. However, for investor inflation uncertainty, macroeconomic predictors exhibit superior accuracy as they consistently outperform the alternative data sources across all horizons, except at one month forecast horizon. Neither of macroeconomic or alternative predictors can offer significant forecast gains when predicting professional forecasters' uncertainty. Examining predictability for households' inflation uncertainty over time reveals that macroeconomic factors were the primary informative source roughly until 2017-2019, after which Google and News Trends became more significant, highlighting the relevance of these emergent data sources during the Covid-19 pandemic and periods of geopolitical tension and supply shocks.

5.2 Weaknesses and Ideas for Future Research

While a rich set of empirical evidence has revealed the significant role of alternative data such as Google Trends and News Trends in forecasting and nowcasting several macroeconomic indicators, a large amount of work could be further developed in the future research in each chapter of this thesis.

First of all, this thesis conducts forecasting and nowcasting exercises using only the latest available data vintages, without taking into account data revisions. Although in most cases employing a real-time dataset is not feasible due to data limitations, ignoring data revisions could distort the actual forecasting power of the predictors and econometric models. This issue is particularly relevant in Chapter 2, where the incorporation of newly revised data points could potentially influence in a material way the evolution of nowcast estimates throughout the reference quarter. Therefore, testing the informational content of these alternative predictors within a real-out-of-sample forecasting framework could enhance our understanding of their role and impact on forecasting accuracy.

Secondly, although this thesis establishes that combining alternative data sources such as Google Trends with various econometric and statistical models can yield significant gains when forecasting macroeconomic variables, it falls short of rigorously explaining the underlying reasons for these improvements. For instance, Chapter 3 demonstrates that random forests exhibits the highest forecast accuracy, raising critical questions about the underlying drivers of this finding: Is this due to its variable selection mechanism, or is it because of its ability to capture nonlinearities? Furthermore, Chapter 4 shows that Google search data are quite effective in forecasting household's uncertainty, but less so for professional forecasters' uncertainty. One reason behind this finding might be that households turn to Google searches to stay informed about economic developments, whereas professionals may not depend as heavily on search engine data. However, this remains a theoretical conjecture, and more rigorous and systematic approaches to answer these questions would be crucial in establishing the definitive usefulness of these data and methods.

Finally, Chapter 4 uses disagreement-based series as a proxy for inflation uncertainty. While several papers utilise similar indicators to measure uncertainty (Hahm and Steigerwald, 1999; Hayford, 2000; Giordani and Söderlind, 2003; Dovern, 2015) there are conflicting findings on whether cross sectional dispersion of point forecasts is a useful proxy for uncertainty. Zarnowitz and Lambros (1987) states that employing cross-sectional dispersion measures as proxies for uncertainty relies on the assumption that these measures are representative of the average dispersion of predictive probabilities held by individual experts, an assumption, however, that cannot be taken for granted.

Therefore, using survey data which can provide a more direct assessment of uncertainty where future economic expectations are gathered as subjective probability distributions, as an alternative measure to disagreement could be an interesting avenue for future research. For example, Söderlind et al. (2011) demonstrates that indicators of inflation uncertainty derived from survey data prove to be effective in assessing inflation risk premia, thereby offering insights into the dynamics of break-even inflation.

Bibliography

Abbas, A. E., Bakır, N. O., Klutke, G.-A., and Sun, Z. (2013). Effects of risk aversion on the value of information in two-action decision problems. *Decision Analysis*, 10(3):257–275.

Altig, D., Baker, S., Barrero, J. M., Bloom, N., Bunn, P., Chen, S., Davis, S. J., Leather, J., Meyer, B., Mihaylov, E., et al. (2020). Economic uncertainty before and during the COVID-19 pandemic. *Journal of Public Economics*, 191:104274.

Ayat, L. and Burridge, P. (2000). Unit root tests in the presence of uncertainty about the non-stochastic trend. *Journal of Econometrics*, 95(1):71–96.

Babii, A., Ghysels, E., and Striaukas, J. (2022). Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*, 40(3):1094–1106.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, 25(1):52–60.

Bai, J. and Ng, S. (2008a). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317.

Bai, J., Ng, S., et al. (2008b). Large dimensional factor analysis. *Foundations and Trends® in Econometrics*, 3(2):89–163.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Bańbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). Now-casting and the real-time data flow. In *Handbook of Economic Forecasting*, volume 2, pages 195–237. Elsevier.

Bantis, E., Clements, M. P., and Urquhart, A. (2023). Forecasting GDP growth rates in the United States and Brazil using Google Trends. *International Journal of Forecasting*, 39(4):1909–1924.

Barbaglia, L., Consoli, S., and Manzan, S. (2023). Forecasting with economic news. *Journal of Business & Economic Statistics*, 41(3):708–719.

Barhoumi, K., Darné, O., and Ferrara, L. (2010). Are disaggregate data useful for factor analysis in forecasting French GDP? *Journal of Forecasting*, 29(1-2):132–144.

Barhoumi, K., Darné, O., and Ferrara, L. (2014). Dynamic factor models: A review of the literature. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2013(2):73–107.

Bernanke, B. S. and Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(3):525–546.

Bilgin, M. H., Demir, E., Gozgor, G., Karabulut, G., and Kaya, H. (2019). A novel index of macroeconomic uncertainty for Turkey based on Google-Trends. *Economics Letters*, 184:108601.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Borup, D. and Schütte, E. C. M. (2022). In search of a job: Forecasting employment growth using Google Trends. *Journal of Business & Economic Statistics*, 40(1):186–200.

Bragoli, D. and Fosten, J. (2018). Nowcasting Indian GDP. *Oxford Bulletin of Economics and Statistics*, 80(2):259–282.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24:123–140.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Bulligan, G., Marcellino, M., and Venditti, F. (2015). Forecasting economic activity with targeted predictors. *International Journal of Forecasting*, 31(1):188–206.

Buono, D., Mazzi, G. L., Kapetanios, G., Marcellino, M., and Papailias, F. (2017). Big data types for macroeconomic nowcasting. *Eurostat Review on National Accounts and Macroeconomic Indicators*, 1(2017):93–145.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Carrière-Swallow, Y. and Labbé, F. (2013). Nowcasting with Google Trends in an emerging market. *Journal of Forecasting*, 32(4):289–298.

Carroll, C. D. (2003). Macroeconomic expectations of households and professional forecasters. *The Quarterly Journal of Economics*, 118(1):269–298.

Carroll, C. D., Fuhrer, J. C., and Wilcox, D. W. (1994). Does consumer sentiment forecast household spending? if so, why? *The American Economic Review*, 84(5):1397–1408.

Caruso, A. (2018). Nowcasting with the help of foreign indicators: The case of Mexico. *Economic Modelling*, 69:160–168.

Castelnuovo, E. and Tran, T. D. (2017). Google it up! a Google trends-based uncertainty index for the United States and Australia. *Economics Letters*, 161:149–153.

Cepni, O., Güney, I. E., and Swanson, N. R. (2019). Nowcasting and forecasting GDP in emerging markets using global financial and macroeconomic diffusion indexes. *International Journal of Forecasting*, 35(2):555–572.

Chen, J. C., Dunn, A., Hood, K., Driessen, A., and Batch, A. (2019). Off to the races: A comparison of machine learning and alternative data for predicting economic indicators. In *Big Data for 21st Century Economic Statistics*. University of Chicago Press.

Cheng, K., Huang, N., and Shi, Z. (2021). Survey-based forecasting: To average or not to average. *Behavioral Predictive Modeling in Economics*, pages 87–104.

Choi, H. and Varian, H. (2009). Predicting initial claims for unemployment benefits. *Google Inc*, 1:1–5.

Claveria, O. (2021). On the aggregation of survey-based economic uncertainty indicators between different agents and across variables. *Journal of Business Cycle Research*, 17(1):1–26.

Clements, M. P. (2016). Real-time factor model forecasting and the effects of instability. *Computational Statistics and Data Analysis*, 100:661–675.

Clements, M. P. and Galvão, A. B. (2019). Data revisions and real-time forecasting. *The Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

Clements, M. P., Rich, R. W., and Tracy, J. S. (2023). Surveys of professionals. In *Handbook of Economic Expectations*, pages 71–106. Elsevier.

Combes, S. and Bortoli, C. (2016). Nowcasting with Google Trends, the more is not always the better. In *CARMA 2016: 1st International Conference on Advanced Research Methods in Analytics*, pages 15–22. Editorial Universitat Politècnica de València.

Croushore, D. (2005). Do consumer-confidence indexes help forecast consumer spending in real time? *The North American Journal of Economics and Finance*, 16(3):435–450.

Dahlhaus, T., Guénette, J. D., and Vasishtha, G. (2017). Nowcasting BRIC+M in real time. *International Journal of Forecasting*, 33(4):915–935.

Diebold, F. X. (2003). Big data dynamic factor models for macroeconomic measurement and forecasting. In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society,"(edited by M. Dewatripont, LP Hansen and S. Turnovsky)*, volume 115, page 22. Citeseer.

Diebold, F. X. (2021). What's the big idea? "Big Data" and its origins. *Significance*, 18(1):36–37.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63.

Donadelli, M. (2015). Google search-based metrics, policy-related uncertainty and macroeconomic conditions. *Applied Economics Letters*, 22(10):801–807.

Doornik, J. A. and Hendry, D. F. (2015). Statistical model selection with "Big Data". *Cogent Economics & Finance*, 3(1):1045216.

Dovern, J. (2015). A multivariate analysis of forecast disagreement: Confronting models of disagreement with survey data. *European Economic Review*, 80:16–35.

Doz, C. and Fuleky, P. (2020). Dynamic factor models. *Macroeconomic forecasting in the era of big data*, pages 27–64.

Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164(1):188–205.

Dzielinski, M. (2012). Measuring economic uncertainty and its impact on the stock market. *Finance Research Letters*, 9(3):167–175.

D'Amuri, F. and Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, 33(4):801–816.

Elliott, G., Gargano, A., and Timmermann, A. (2013). Complete subset regressions. *Journal of Econometrics*, 177(2):357–373.

Ferrara, L. and Simoni, A. (2019). When are Google data useful to nowcast GDP? an approach via pre-selection and shrinkage.

Ferrara, L. and Simoni, A. (2023). When are Google data useful to nowcast GDP? an approach via preselection and shrinkage. *Journal of Business & Economic Statistics*, 41(4):1188–1202.

Foroni, C. and Marcellino, M. (2014). A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates. *International Journal of Forecasting*, 30(3):554–568.

Foroni, C., Marcellino, M., and Stevanovic, D. (2022). Forecasting the Covid-19 recession and recovery: Lessons from the financial crisis. *International Journal of Forecasting*, 38(2):596–612.

Garcia, M. G., Medeiros, M. C., and Vasconcelos, G. F. (2017). Real-time inflation forecasting with high-dimensional models: The case of Brazil. *International Journal of Forecasting*, 33(3):679–693.

Giacomini, R. and Granger, C. W. (2004). Aggregation of space-time processes. *Journal of Econometrics*, 118(1-2):7–26.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

Giordani, P. and Söderlind, P. (2003). Inflation forecast uncertainty. *European Economic Review*, 47(6):1037–1059.

Gogas, P., Papadimitriou, T., and Agrapetidou, A. (2018). Forecasting bank failures and stress testing: A machine learning approach. *International Journal of Forecasting*, 34(3):440–455.

Götz, T. B. and Knetsch, T. A. (2019). Google data in bridge equation models for German GDP. *International Journal of Forecasting*, 35(1):45–66.

Hahm, J.-H. and Steigerwald, D. G. (1999). Consumption adjustment under time-varying income uncertainty. *Review of Economics and Statistics*, 81(1):32–40.

Hayford, M. D. (2000). Inflation uncertainty, unemployment uncertainty and economic activity. *Journal of Macroeconomics*, 22(2):315–329.

Hendry, D. F. and Hubrich, K. (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of Business & Economic Statistics*, 29(2):216–227.

Hubrich, K. (2005). Forecasting Euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy? *International Journal of Forecasting*, 21(1):119–136.

Inoue, A. and Kilian, L. (2008). How useful is bagging in forecasting economic time series? A case study of US consumer price inflation. *Journal of the American Statistical Association*, 103(482):511–522.

Jansen, W. J., Jin, X., and de Winter, J. (2016). Forecasting and nowcasting real GDP: Comparing statistical models and subjective forecasts. *International Journal of Forecasting*, 32(2):411–436.

Kalamara, E., Turrell, A., Kapetanios, G., Kapadia, S., and Redl, C. (2020). Making text count for macroeconomics: What newspaper text can tell us about sentiment and uncertainty. In *2019 European Economic Association Meeting*.

Kapetanios, G., Papailias, F., et al. (2018). Big data & macroeconomic nowcasting: Methodological review. *Economic Statistics Centre of Excellence, National Institute of Economic and Social Research*.

Keane, M. P. and Runkle, D. E. (1990). Testing the rationality of price forecasts: New evidence from panel data. *The American Economic Review*, pages 714–735.

Kim, H. H. and Swanson, N. R. (2018). Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting*, 34(2):339–354.

Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics*, 186(2):325–344.

Kohlscheen, E. (2022). What does machine learning say about the drivers of inflation? *arXiv preprint arXiv:2208.14653*.

Kotchoni, R., Leroux, M., and Stevanovic, D. (2019). Macroeconomic forecast accuracy in a data-rich environment. *Journal of Applied Econometrics*, 34(7):1050–1072.

Lemieux, J. and Peterson, R. A. (2011). Purchase deadline as a moderator of the effects of price uncertainty on search duration. *Journal of Economic Psychology*, 32(1):33–44.

Luciani, M., Pundit, M., Ramayandi, A., and Veronese, G. (2018). Nowcasting Indonesia. *Empirical Economics*, 55(2):597–619.

Ludvigson, S. and Bram, J. (1998). Does consumer confidence forecast household expenditure?: A sentiment index horse race. *FRBNY Economic Policy Review*, 4(2):59–78.

Lütkepohl, H. (2006). Forecasting with VARMA models. *Handbook of Economic Forecasting*, 1:287–325.

Manski, C. F. (2004). Measuring expectations. *Econometrica*, 72(5):1329–1376.

Marcellino, M., Stock, J. H., and Watson, M. W. (2003). Macroeconomic forecasting in the Euro area: Country specific versus area-wide information. *European Economic Review*, 47(1):1–18.

Mariano, R. S. and Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18(4):427–443.

McCracken, M. W. and Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Medeiros, M. C. and Mendes, E. F. (2016). L1-regularization of high-dimensional time-series models with non-gaussian and heteroskedastic errors. *Journal of Econometrics*, 191(1):255–271.

Medeiros, M. C. and Pires, H. F. (2021). The proper use of Google Trends in forecasting models. *arXiv preprint arXiv:2104.03065*.

Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., and Zilberman, E. (2021). Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business & Economic Statistics*, 39(1):98–119.

Mishkin, F. S. (2008). Awhither federal reserve communication,@ speech delivered at the peterson institute for international economics. *Washington, DC,(July 28)(www. federalreserve. gov/newsevents/speech/mishkin20080728a. htm)*.

Møller, S. V., Pedersen, T., Montes Schütte, E. C., and Timmermann, A. (2023). Search and predictability of prices in the housing market. *Management Science*.

Newey, W. K., West, K. D., et al. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Panagiotelis, A., Athanasopoulos, G., Hyndman, R. J., Jiang, B., and Vahid, F. (2019). Macroeconomic forecasting for Australia using a large number of predictors. *International Journal of Forecasting*, 35(2):616–633.

Politis, D. N. and Romano, J. P. (1991). *A circular block-resampling procedure for stationary data*. Purdue University. Department of Statistics.

Politis, D. N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70.

Rapach, D. E. and Strauss, J. K. (2012). Forecasting US state-level employment growth: An amalgamation approach. *International Journal of Forecasting*, 28(2):315–327.

Ravn, M. O. and Uhlig, H. (2002). On adjusting the Hodrick-Prescott filter for the frequency of observations. *Review of Economics and Statistics*, 84(2):371–376.

Schiavoni, C., Palm, F., Smeekes, S., and van den Brakel, J. (2021). A dynamic factor model approach to incorporate big data in state space models for official statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184(1):324–353.

Schumacher, C. (2010). Factor forecasting using international targeted predictors: The case of German GDP. *Economics Letters*, 107(2):95–98.

Seabold, S. and Coppola, A. (2015). Nowcasting prices using Google trends: an application to Central America. *World Bank Policy Research Working Paper*, (7398).

Shapiro, A. H., Sudhof, M., and Wilson, D. J. (2022). Measuring news sentiment. *Journal of Econometrics*, 228(2):221–243.

Smeekes, S. and Wijler, E. (2018). Macroeconomic forecasting using penalized regression methods. *International Journal of Forecasting*, 34(3):408–430.

Söderlind, P. et al. (2011). Inflation risk premia and survey evidence on macroeconomic uncertainty. *International Journal of Central Banking*, 7(2):113–133.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Stock, J. H. and Watson, M. W. (2011). Dynamic factor models. *Oxford Handbooks Online*.

Thorsrud, L. A. (2020). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics*, 38(2):393–409.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Vosen, S. and Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting*, 30(6):565–578.

Woloszko, N. (2020). Tracking activity in real time with Google Trends. Technical report, OECD Publishing.

Woo, J. and Owen, A. L. (2019). Forecasting private consumption with Google Trends data. *Journal of Forecasting*, 38(2):81–91.

Yu, L., Zhao, Y., Tang, L., and Yang, Z. (2019). Online big data-driven oil consumption forecasting with Google trends. *International Journal of Forecasting*, 35(1):213–223.

Zarnowitz, V. and Lambros, L. A. (1987). Consensus and uncertainty in economic prediction. *Journal of Political Economy*, 95(3):591–621.

Zou, H. (2006). The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (statistical methodology)*, 67(2):301–320.

Zou, H., Hastie, T., Tibshirani, R., et al. (2007). On the "degrees of freedom" of the LASSO. *The Annals of Statistics*, 35(5):2173–2192.