Norwegian School of Economics
Lisbon, Spring 2022

Searching for Inflation

An Empirical Study of Real-Time Micro Behaviour Data on Inflation Nowcasts

Siem Yemane Haile and Lars Berg Strømmen

Supervisor: Krisztina Molnár

Master thesis, Economics and Business Administration

Major: Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

This thesis is written as a part of our Master of Science degree in Business and Administration at the Norwegian School of Economics (NHH).

During our time at NHH, we have developed a strong interest in courses at the intersection of economics and financial economics. The topic for this thesis reflects this personal and academic interest. We especially found the courses *International Finance*, *Business Cycle Analysis*, *Money and Banking*, and *Macroeconomic Theory and Policy* attractive, and choosing a topic within this field of academia felt natural.

Writing this thesis started in the spring of 2021 and lasted for over a year. This long period has both been comfortable and cumbersome as there have always been other assignments and tests of a more pressing issue. Procrastination has been a constant temptation along the way. In addition, writing the thesis remotely from two countries, with non-overlapping time schedules, has been an exciting challenge. However, we are satisfied with the result and with the accumulated knowledge on an interesting topic we have acquired along the way.

We want to take the opportunity to thank our much supporting and understanding supervisor, Krisztina, for being most helpful and facilitating us working remotely. The feedback and support have been of great assistance. Secondly, we would like to thank Inga Nielsen Friis at Norges Bank for significant input along the way. Finally, thanks go out to friends and family who have listened patiently and understandingly to rants of frustration on a topic they have no interest in and for the motivating words along the way.

Norwegian School of Economics

Lisbon, May 2022

Siem Yemane Haile

Lars Berg Strømmen

Abstract

This thesis introduces the application of real-time micro behaviour data in inflation nowcasts. Our study analyses if ARIMA models extended with Google search data improves the prediction of divisions of inflation compared to the high-performing simple AR(1) process. This analysis addresses the issue of official inflation data containing a lag of ten days. Real-time micro behaviour data can contain valuable information, which provides policymakers with a new tool to predict inflation in the present and near future.

First, each division of inflation is assigned corresponding Google Indicators before in-sample model selection is performed using the Box-Jenkins Methodology. Then, comparisons against ARIMA baselines are conducted to evaluate if Google search data improve model selection. Further, out-of-sample predictions are performed for the improved divisions from the preceding step. Finally, the nowcast performance for each division is compared against the simple AR(1) process in terms of prediction error and ability to identify trends and turning points.

This thesis documents that Google search data improves model selection for six of twelve divisions of inflation. These divisions consist of goods and are volatile compared to the remaining six. Furthermore, four of six extended ARIMA models outperform the simple AR(1) process in prediction error for the out-of-sample nowcasts. At the same time, all divisions are improved in predicting trends and turning points. These findings suggest that real-time micro behaviour data, represented by Google Trends, improve model selection and nowcasts of some divisions compared to AR(1). However, when compared to replicated and baseline ARIMA models, the only value of Google search data is in model selection. The improved performance is attributed to the properties of ARIMA. To conclude, real-time data on micro behaviour are of value in model selection in inflation nowcasts.

Keywords – Google Trends, Inflation, Nowcast, Macroeconomic Modeling

Contents

List of Figures

List of Tables

1 Introduction

Accurate and timely forecasts of macroeconomics indicators are imperative for central banks and governments to implement prompt and effective policy decisions. One of the most important indicators is the Consumer Price Index (CPI) which measures the inflation level by indicating the rise in consumer prices of commodities and services (SSB, 2022). CPI provides an essential foundation for the decision-making of monetary policy, fiscal policy, and national economic accounting. Improper or inaccurate forecasts of CPI can consequently lead to an impediment to economic growth. This thesis aims to investigate if inflation nowcasts can benefit from incorporating real-time data on micro behaviour represented by Google Trends.

Norwegian core inflation data are branched into twelve divisions of different goods and services (SSB, 2022). The classification criteria are based on the end purpose of the consumption, known as COICOP.[1] This composition of consumer prices facilitates understanding of where changes in inflation originate from, thus enabling the identification of which variables drive the change. One approach to such analysis is to incorporate customised data on micro behaviour related to specific divisions of CPI as explanatory variables.

Inflation data has historically been incomplete as it is released with a lag. This lag creates difficulties in understanding the real-time changes in inflation (Giannone et al., 2006). For example, data regarding inflation in Norway are usually released ten days into the following month (SSB, 2022). This delay, together with incomplete models, may lead to a recession if inappropriate policies are implemented in the meantime.

Historically, the top-performing models have not utilised real-time data to predict inflation. One such model is the simple autoregressive process of order 1 (AR(1)) (Chan, 2011). This model uses lagged values of inflation to predict the future. The strength and weakness of this model have been its simple nature. On the one hand, predicting based on lagged observations explains something about recent developments. On the other hand, it is not a guarantee of what the future holds. This thesis investigates whether incorporating real-time data on micro behaviour improves the predictive ability of the AR(1).

[1] Classification of Individual Consumption by Purpose

In economics, the prediction of the present, near future, and the recent past state is known as nowcasting (Stock & Watson, 2008). The real-time data used in nowcasting may provide information about the present which can support decision-makers in making more effective and precise policies. Nowcasting originates from meteorology, which has long been used, while its application in economics has proliferated in recent years (Bańbura et al., 2013). For example, GDP is an indicator that has been nowcast numerous times in Norway. However, efforts to nowcast inflation are of limited attempts. This thesis aims to supplement the scarce literature on nowcasting inflation by utilising real-time search data in Norway.

Google Trends is a public tool that supplies data on online search behaviour. The user-generated data can reflect users' opinions and public sentiment on economic issues (Choi & Varian, 2012). For instance, if the public wants to acquire knowledge on some topic, they usually use a search engine to extract information represented by some keywords. If consumers notice that housing prices increase rapidly, they may search for some combination of words such as *"housing price"* or *"rise in housing prices"*. An increase in the relative number of searches related to these terms may represent the consumer sentiment on price development at any given moment. Moreover, Google Trends allows for filtering keywords by specific categories. This filtering facilitates analysing developments within specific areas.

The issue of models predicting based on lagged inflation values may be solved by incorporating real-time data representing micro behaviour. Furthermore, as Google Trends allows for isolating keywords for specific categories, we pursue a method that connects search data to divisions of inflation. This method facilitates granular predictions. Thus, the following research question is proposed:

Can micro behaviour data represented by Google search queries predict divisions of inflation in real-time more precisely than the simple AR(1) process?

The research question is answered through the three sub-questions presented below:

Can Google search queries improve in-sample model selection for divisions of inflation relative to baseline models from 2011 to 2020?

Can Google search queries reduce out-of-sample prediction error to CPI ex-post through 2021 compared to the AR(1) model?

Can Google search queries improve the out-of-sample ability to predict short-term changes in inflation through 2021 compared to the AR(1) model?

This thesis follows a sequential process to arrive at the research question. In the first step, we conduct a preliminary test of content by constructing a set of AR(1) models for each division of inflation. These represent the baseline AR(1) models. We then extend our baseline models by adding Google Indicators to investigate if Google search queries are information value-adding. We aim to identify if these extended models reduce the RMSE compared to their respective baselines.

The next step extends the analysis by using the acknowledged Box-Jenkins methodology to identify ARIMA models for each division of inflation. Models are ranked and selected by the estimator of prediction error they yield through the in-sample period from January 2011 to December 2020.

Lastly, based on Step 2, we proceed with the divisions where extended models were identified as a better fit for prediction than their respective baselines. Subsequently, we perform an out-of-sample one-month rolling nowcast through 2021 to determine the predictive ability of Google data. In addition, we compare our extended ARIMA models to AR(1) models before benchmarking both against CPI ex-post.

The thesis is divided into a broad structure of eight chapters. Chapter 2 discusses why accurate forecasts of inflation are essential for decision-makers and how real-time data can improve current models for predicting inflation. In addition, the chapter reviews the existing body of literature on predicting inflation, focusing mainly on papers employing Google search data. In Chapter 3, we describe the collected data. In addition, the selection process and which search terms were selected as Google Indicators are presented. Chapters 4 and 5 review the models and methodology used to predict inflation, while Chapters 6 and 7 present and discuss the nowcasts' results. Finally, Chapter 8 concludes this paper.

2 Background

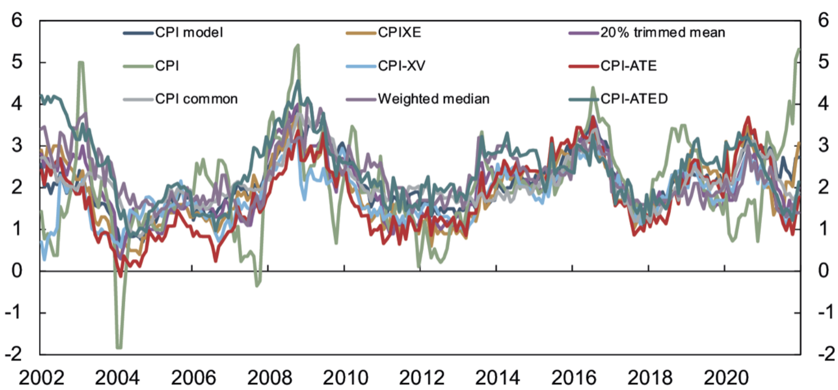
2.1 Relevance

Low and stable inflation is the primary target of Norges Bank (Olsen, 2022). Sustained high inflation comes at a great cost for society. The value of money becomes uncertain, and households and firms find it increasingly cumbersome to plan. Thus, ensuring stability in consumer prices is at the core of Norges Bank's social mission.

Nonetheless, central banks appreciate a certain level of inflation, as increasing prices contribute to flexibility in the economy. This is the foundation of Norges Bank's flexible inflation targeting. Today, the aim is for the increase in annual consumer prices to remain close to 2% over time.[2] Nevertheless, Norges Bank is not unequivocally concerned with navigating inflation towards its long-term target. Credibility and confidence that the central bank is determined to achieve low and stable inflation over time offer Norges Bank flexibility in allowing for fluctuations around the target at a low socio-economic cost.

The purpose of indicators for underlying inflation is to control for transitory volatility in consumer prices (Norges Bank, 2022). Some price components of the broad index, such as prices of energy products, are prone to large fluctuations from one period to the next. Thus, price volatility for such components creates noise to the underlying trend. As such, Norges Bank uses a range of indicators to provide measurements of the real-time underlying trend in consumer prices (see Figure 2.1).

Figure 2.1: Norges Bank's indicators for consumer prices



[2]On 2nd March 2018, Norges Bank announced their new target of 2%. The central bank had operated with a target of 2.5% +/-1 pp for 17 years before aligning the target with its trading partners.

In particular, Norges Bank closely monitors the Consumer Price Index adjusted for tax changes and excluding energy products (CPI-ATE). This indicator of consumer prices is a measurement of underlying growth in CPI, known as *core inflation*. The inflation target of Norges Bank relates specifically to this indicator.

Accurate and timely information is of great importance in implementing policies that support the central bank's objectives. Nowcasting provides a tool for tracking inflation in real-time. In addition, it facilitates understanding the underlying forces that may endanger price stability while they are ongoing. Thus, it can support a central bank in identifying when to act and what actions to take to offset these forces with timely and effective policies (Modugno, 2013).

Extending this toolbox with real-time micro behaviour data allows policymakers to study current inflation at more granular levels. To date, nowcasting inflation with Google Trends is something that has not been done in Norway. This master thesis aims to fill this void by exploring the predictive ability of Google search data for real-time predictions of inflation divisions.

2.2 Literature Review

This chapter outlines the related literature to this thesis. It is divided into three main parts. First, literature on predicting inflation is outlined before previous research on nowcasting is discussed. Further, existing literature related to nowcasting inflation is presented. Finally, criticism of using Google Trends for nowcasting is discussed to conclude this section.

2.2.1 Prediction of Inflation

Forecasting the future rate of inflation has long been of interest in the academic world as this macroeconomic phenomenon has historically been important for the economic development of countries (Stock & Watson, 2008). Standard forecasting models apply macroeconomic indicators as independent variables to predict future CPI changes. Economic growth, investment, money supply, and resident income have generally been used in such models. In addition to these indicators, the past information of the dependent and independent variables is included. Historically, autoregressive models (AR), autoregressive distributed lag models (ADL), autoregressive integrated moving average models (ARIMA), error correction models (ECM), and vector autoregressive models (VAR) are the most used time series models for prediction (Li, Shang, Shouyang, & Ma, 2015). In addition, econometric, artificial intelligence, and regime-switching models have been used to forecast inflation (Ang, Bekaert, & Wei, 2007; Hamilton, 1996; Rousseau & Wachtel, 2002). Later, these have been combined into an integrated model that uses different techniques to improve the forecasts (Choudhary & Haider, 2012; Lai, Hsieh, & Chang, 2005; Nakamura, 2005; Svensson, 2000).

The models mentioned above generally use data that are not real-time. The time series models are generally forecast using the time series of CPI itself or statistically correlated economic indicators to predict inflation (Li, Shang, Shouyang, & Ma, 2015). They all rely on published statistical data and rarely use real-time data which considers consumer behaviour. Data representing micro behaviour and consumer sentiment should be incorporated to improve inflation forecasts.

2.2.2 Nowcasting Using Google Trends

Google search data reflect public opinion in real-time for specific terms (Li, Shang, Shouyang, & Ma, 2015). Furthermore, it is constructed through a time sequence; thus, it is easy to measure and manipulate compared to other data generated on the internet. As a result of the properties of Google Trends, a growing literature has provided evidence of this data's usefulness in forecasting the economy's current state. For instance, Ettredge et al. (2005), Kholodilin et al. (2009), Choi & Varian (2009) (2012), Guzmán (2011), Carrière-Swallow & Labbé (2013), Chen et al. (2015), Narita & Yin (2018), Ferrara & Simoni (2019), and Woloszko (2020) have utilised Google search data to track and nowcast various economic variables and activities.

Other papers have researched the connection between Google Trends and unemployment. D'Amuri et al. (2017), Fondeur & Karamé (2013), and Baker & Fradkin (2017) have all researched the topic, while Pisu, Costa, & Hwang (2020) nowcasted the state of digitisation. One of the most prominent topics of nowcasting using Google Trends is housing prices. Askitas & Zimmermann (2009) and Wu & Brynjolfsson (2015) have provided theoretical work in this area. More recently, Abay et al. (2020) and Doerr & Gambacorta (2020) used

Google Trends data to assess the impact of the Covid-19 crisis.

One of the key issues when creating nowcasting models using Google Trends is the choice of keywords. Two different approaches have been applied in the existing literature. The first method is illustrated in Da et al. (2011), where specific keywords were theoretically linked to the dependent variable. Goel et al. (2010) pursue another approach where Google Categories are used as the basis for choosing keywords to ensure that the search queries optimally describe the dependent variable of interest.

Results from existing literature on using Google Trends data as a proxy for consumer sentiment are promising. This fact creates an option for the survey-based sentiment data, which have traditionally been used in creating and forecasting leading macroeconomic indicators. Vosen & Schmidt (2011) illustrate in their paper that Google Trends outperforms the Conference Board Consumer Confidence Index and the University of Michigan Consumer Sentiment Index in predicting private consumption in the United States. Another study that confirms the same result is Carrière-Swallow & Labbé (2013), where an index of consumer interest in purchasing automobiles was developed. This index outperformed the benchmark, IMACEC. These results confirm that Google Trends can represent public sentiment and that using this data to capture micro-behaviour to predict real-time changes is feasible.

2.2.3 Nowcasting Inflation Using Google Trends

Regarding the nowcasting methodology, we use two papers as key literature; Forecasting Inflation by Stock & Watson (2008) and *Predicting the Present with Google Trends* (2012) by Varian & Choi. This thesis relates closely to these papers. Varian & Choi supplement AR(1) processes with Google indicators to analyse if prediction error (RMSE) is reduced. Furthermore, we apply the same approach for nowcasting inflation as Stock and Watson. This approach uses an *in-sample period* to estimate and select models as a basis for predicting and an *out-of-sample period* to evaluate the performance of the nowcasts. These papers are recognised and peer-reviewed, and we use these as a basis for our thesis.

Our research is closely related to the empirical work done by Seabold & Coppola (2015) and Li, Shang, Shouyang, & Ma (2015) on how internet search keyword data can be added to predicting inflation. For example, Seabold & Coppola (2015) created an index of Google search queries related to inflation in Central America and estimated one-step-ahead forecasts for several price indices for consumer goods. Further, they compare the performances of the nowcasts to models which historically have performed well.

Li, Shang, Shouyang, & Ma (2015) nowcast inflation in China by using Google Trends in MIDAS models.[3] Both papers show that models extended with Google Indicators can improve nowcasts of inflation compared to AR(1) and ARIMA models. Therefore, we will use a similar framework as these two models in nowcasting inflation.

We supplement the scarce literature by nowcasting inflation for all divisions. Existing literature is limited to nowcasting the aggregated level of inflation or some specific divisions. In addition, our method for choosing keywords is distinct from previous research. We use a combination of the approach of Da et al. (2011), Goel et al., and Li, Shang, Shouyang, & Ma (2015). This approach ensures that our keyword and Google category selection coincide with each division of inflation.

2.2.4 Criticism of the Use of Google Search Queries for Prediction

Google Trends are often subject to limitations as their primary purpose is, in general, not scientific analysis (Woloszko, 2020). For example, in the Google Trends data, structural breaks in January 2011 and January 2016 are caused by changes in the data collection process. In addition, the number of users has rapidly increased since the introduction of the tool in 2004. This increase leads to the relative search intensities of most queries decreasing over time as the time series become less volatile. All these caveats require specific action and statistical processing to avoid affecting the validity of the research.

Another criticism of using Google search queries is that keywords may have alternate meanings. Therefore, it is difficult to separate which meaning of the keyword the search activity relates to (Samanta, 2019). For instance, "price" can also mean "award" in Norwegian. This double-meaning will lead to the model overestimating the actual search volume of the series.

There have been instances of forecasting models using Google Trends not being valid across time. Lazer et al. (2014) argue that models based on Big Data and Google Trends can go from being valid and high performing to producing substandard results. The

[3] Mixed-Data Sampling

reasoning behind this drastic decline in forecasting performance is attributed to the ability of the Google algorithm to associate keywords correctly. The paper further points out that the algorithm may incorrectly include multiple keywords in topics and categories which do not relate to the independent variable.

Lastly, Lazer et al. (2014) point out that unusual attention regarding some keywords may affect the search activity, especially for search queries with low search volumes. News articles mentioning the keywords in question might increase the search activity extensively, which does not represent the actual change in users' sentiment towards the keyword in question.

3 Data

This chapter is dedicated to describing the data collected. First, we present the relevant data, namely the consumer price index and Google Trends. Next, we discuss the properties of the data before presenting the selection criteria for selection of keywords. Finally, the collected data are listed. The perspective taken in this thesis ranges from January 2011 to December 2021. This period is split into two samples. Throughout the thesis, the former will be referred to as the in-sample subset, starting January 2011 and ending December 2020. The latter is the out-of-sample subset, ranging from January 2021 to December 2021.

3.1 Consumer Price Index

The Consumer Price Index (CPI) is a measurement that describes the development in consumer prices for goods and services (SSB, 2022). Specifically, this broad indicator describes the development in the price of a weighted average market basket of consumer items purchased by firms and households. This thesis emphasises the twelve-month growth rate (YOY) of consumer prices in this thesis.

Norges Bank is arguably the institution with the greatest desire to measure and predict inflation at any given moment (see Section 2.1). However, the central Norwegian office for official government statistics, Statistics Norway (SSB), measures and publishes the consumer prices in Norway. The data are published monthly, at monthly frequencies, ten days trailing to the month in question. Thus, data on consumer prices exhibit a lag.

Statistics Norway measures consumer prices across five indicators (see Table 3.1). These indicators either include prices of all items or exclude certain types of goods or services that may distort the picture of underlying inflation. The most common items to exclude are taxes, energy products, and electricity. It is beyond the scope of this thesis to go into details about all indicators, as the purpose is to improve decision-making for policymakers. Thus, our focus is narrowed to CPI-ATE.

Table 3.1: CPI five-level hierarchy of consumption groups

	CPI All-item	CPI-AT	CPI-ATE	CPI-AE	CPI-AEL
Level 1	Divisions	Divisions	Divisions		
Level 2	Groups				
Level 3	Sub-groups 1				
Level 4	Sub-groups 2				
Level 5	Items and items groups				

Further, CPI indicators are deconstructed in up to five consumption groups. Only the broad CPI indicator is deconstructed into all five levels (see Table 3.1). The indicator of interest, CPI-ATE, is available at the division level. This availability facilitates analysing changes in inflation on a more granular level and for specific goods and services. For the remainder of the thesis, the division level of core inflation will be the focus.

3.1.1 Composition of CPI

The index for core inflation is composed of twelve divisions. These divisions are formally known as the European Classification of Individual Consumption According to Purpose (COICOP). Such classification aims to lay down a framework of homogeneous categories of goods and services considered a function or purpose of household expenditures (Statistical Office of the European Communities, 2010). The objective is to classify transactions to individuals' and households' real consumption. In the context of inflation, COICOP establishes weights and aggregated price levels for each division (SOEC, 2010). The divisions of core inflation are listed below (see Table 3.2).

Some of the 12 divisions of core inflation consist solely of goods, while others consist of services or a combination of both (see Table 3.2). Goods are further classified as non-durable, semi-durable, or durable. This classification of the different divisions allows us to investigate if there are some categories of inflation and types of consumption that benefit more than others from including search data to nowcast inflation.

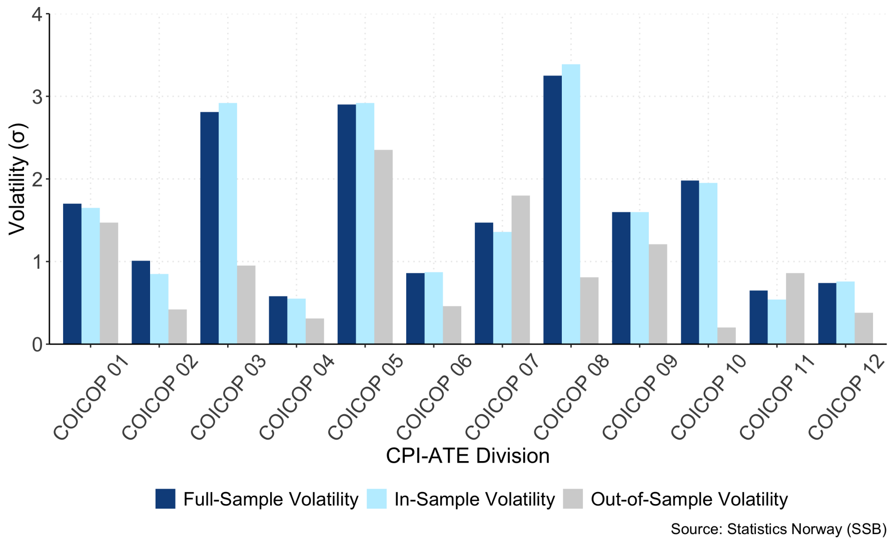
Table 3.2: Classification of CPI divisions. Dark blue indicates majority of consumption type for the division. Light blue the minority.

	Goods			
Division	Durable	Semi-durable	Non-durable	Services
01 Food and non-alcoholic beverages			■	
02 Alcoholic beverages and tobacco			■	□
03 Clothing and footwear		■		
04 Housing, water, electricity, gas and other fuels			□	■
05 Furnishings, household equipment and routine maintenance	■	□		□
06 Health			□	■
07 Transport	■	□	□	□
08 Communications	□			■
09 Recreation and culture	□	□	□	■
10 Education				■
11 Restaurants and hotels				■
12 Miscellaneous goods and services	□	□	□	■

3.1.2 Volatility

Divisions vary in the items they measure and therefore differ in volatility over time (see Figure 3.1). The illustration below makes a distinction between the full sample period (2011 – 2021), in-sample period (2011 – 2020), and out-of-sample period (2021). It is apparent that some divisions are persistently volatile in-sample and remain volatile out-of-sample. Other divisions tend to be considerably less volatile out-of-sample. Overall, divisions appear to have low volatility out-of-sample.

Figure 3.1: Volatility of divisions of inflation full sample, in-sample and out-of-sample



Ten of the twelve divisions stand out for various reasons. First, COICOP 03, 05, 08, and 10 drop drastically from one sample period to the next. The three former were the most volatile in-sample. Second, COICOP 07 and 11 appear as the only divisions that increase in volatility out-of-sample. Both tend to be more volatile than through the entire in-sample period. Third, COICOP 02, 04, 06, 11, and 12 show no more than slight tendencies of volatility through all three sample periods. It is beyond the scope of this thesis to go into detail about the cause of this variation.

3.1.3 Data Collected

Data on inflation were collected from Statistics Norway. We extract twelve monthly data sets for all divisions (see Table 3.2). These are a series of each division of core inflation from January 2011 to December 2021.

3.2 Google Trends

3.2.1 General Description

Google Trends was launched in 2006 to provide free access to data that reflect user interest by comparing the search volume of keywords across time (Combes & Bortoli, 2016). The data is presented in queries or a set of search terms linked semantically. Every search query is scaled from 0 to 100 to the search volume for the chosen period and is available at daily, weekly, monthly, or yearly frequencies.[4] For instance, if the monthly frequency is applied, then the statistic will present the popularity of the keyword for each month to the month with the highest search volume. Conversely, the value is set to zero if the keyword has a relatively insignificant search volume for one period.

Google Trends eliminates duplicated searches before it is published. Therefore, if a person searches for a word repeatedly over a short period, it will only be registered as one search. This feature strengthens the validity of Google Trends as individuals cannot manipulate the data. Thus, Google Trends presents representative data of interests and sentiment.

[4]The availability of frequencies depends on the duration of the window. Monthly frequencies are available from 2004, weekly frequencies for the past five years, daily frequencies for the past 90 days, hourly frequencies for the past seven days, and minutely frequencies for the past 24 hours. Note that the availability of frequencies has no retroactive effect, i.e., hourly frequencies are only available for the past seven days from the current time and not for the first seven days of 2004.

A complimentary tool to Google Trends is Google Adwords, which presents statistics on keywords. This tool facilitates investigating the statistical properties of keywords to ensure validity in the search queries from Google Trends. One limitation of Google Trends is that the absolute number of searches is unavailable. Google Adwords solves this issue as the absolute number of searches for specific keywords are provided.

3.2.2 Normalisation

The Google data are pre-treated. This treatment means that the absolute number of searches for keywords is not made available to the public (Combes & Bortoli, 2016). As the Google search engine processes billions of searches daily, generating the entire data set would be too large of a task to accomplish quickly.

To circumvent this problem, Google indexes the data. This process is formally known as normalisation, which makes it possible to get and process a data set generally representative of all Google searches within minutes of an event. The data are normalised, so the maximum of the search query always equals 100. First, each data point is divided by the total number of searches for the period and location (see Equation 3.1):

$$\text{Normalised value} = \frac{\text{Actual search term volume}}{\text{Total search volume}} \tag{3.1}$$

Second, residual numbers are scaled on a range from 0 to 100 by dividing the normalised value by the highest normalised value (see Equation 3.2):

$$\text{Scale} = \text{Google Index} = \frac{\text{Normalised value}}{\text{Highest normalised value}} * 100 \tag{3.2}$$

As a result, the data from Google Trends represents relative popularity. This normalisation solves the issue where the location with the highest search volume would consistently be ranked the highest. If the values were absolute, areas with high density would dominate less dense areas, which does not indicate anything about the underlying trend. The core of Google Trends is to identify the propensity of individuals to search for different terms and topics on Google on a regular basis.

Lastly, normalisation presents some challenges as each query of the same keyword is different (Woloszko, 2020). The time series might be revised from one date to the next as only samples of the absolute number of searches are made available. The longer time between the two points of extraction, the more significant the discrepancy is between the two series. Thus, direct comparison between two supposedly identical series should be cautiously approached. This discrepancy may threaten the data's validity as it is not stable across time.

To circumvent the issue of different time series, we have to take some considerations. First, the variation in the samples of terms with a higher number of searches does not vary much (Medeiros & Pires, 2021). For instance, the keyword *"price"* is relatively stable for every sample. The number of absolute searches can be checked by using Google Adwords. However, downloaded queries should be collected in multiple samples and averaged across every term. We collect data sets for each Google Indicator from multiple days and average them across time.

3.2.3 Keywords, Topics, and Categories

Google Trends has two keyword alternatives: search terms and topics. Search terms are specific to the inserted keywords, while topics capture related search terms to the inserted keyword. Thus, topics are a collection of multiple keywords.

Further, Google Trends provides a filter for categories. The filter applies to both search terms and topics. In total, there are 1,200 categories. This wide range of categories is due to the probabilistic algorithm applied, which allocates individual keywords into multiple categories (Woloszko, 2020). Moreover, each category is a composite of a 5-level hierarchy. Therefore, filtering a search term or Topic by category returns the result for the keywords compatible with the selected category.

For instance, searching for *"inflation"* with no category returns generic searches on inflation. However, by using the category *"Autos & Vehicles"*, the output is limited to inflation related to autos and vehicles.

Benefits and Challenges of Keywords, Topics, and Categories

The topic feature makes it possible to capture the overall interest in keywords. This feature has several benefits. First, it enables the user of Google Trends to capture other searches that include the keyword in question. Second, topics control for errors such as misspellings, as it includes multiple spellings of a keyword. Third, it addresses the issue of ambiguous keywords such as *"Apple"*, which is both a company and a fruit.

The category feature ensures that keywords' meaning is consistent across time. For instance, searching for *"DNB"* may have reflected the demand for a loan in the early years of 2000. However, in the past ten years, most Norwegians have started to use online banking services. Thus, the intention of searching for *"DNB"* might have changed. In comparison, the meaning of categories and topics do seem more stable to analyse across time compared to keywords.

Despite the benefits of topics and categories, there are limitations. Google does not provide any information on how topics or keywords are constructed. As a result, the selected Topic or filtered category may not represent the user's intended purpose, as one cannot verify which keywords are allocated to these groups.

3.3 Choice of Google Search Data

Google search data are subject to limitations as the primary purpose is not for scientific analysis. Consequently, some considerations and transformations must be taken when applying this data. This section describes the collected data and which considerations have been made. First, we describe the time series length before discussing which frequency should be applied. Further, we discuss in depth which keywords and Google Categories we have selected and why these can be applied to predict inflation.

3.3.1 Length of Time Series

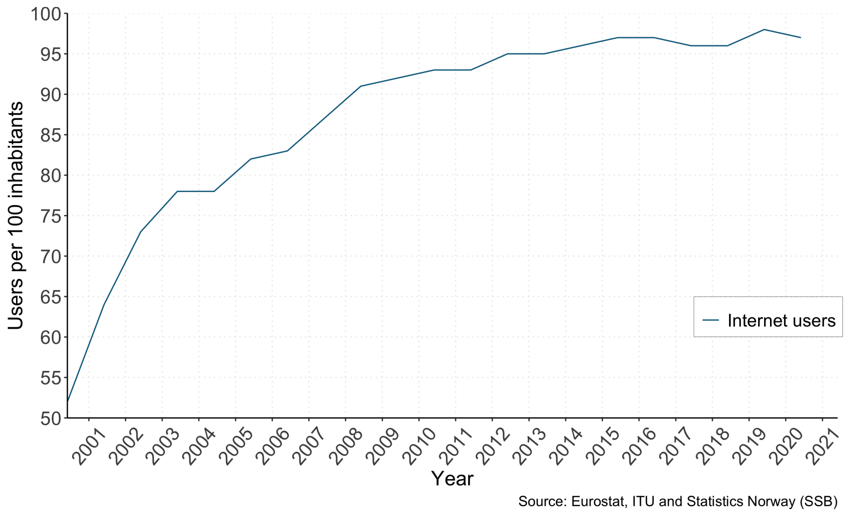
To ensure that we collect a time series appropriate for nowcasting, we will perform several assessments of the Google Trends data. We consider internet penetration, Google's search engine market share, significant events, structural breaks, and the number of observations. Based on these considerations, we will determine if we can use Google Trends to predict inflation and which period will be applied in our models.

Internet Penetration in Norway

Norway has been at the forefront of adopting technology, and internet users have increased rapidly since 2000 (see Figure 3.2). At the start of the millennium, about 50% of the

Norwegian population had access to and used the internet. This number increased to about 90% around 2010 and stabilised around 95% few years later. This stability makes the time series less prone to trends and ensures a high number of observations.

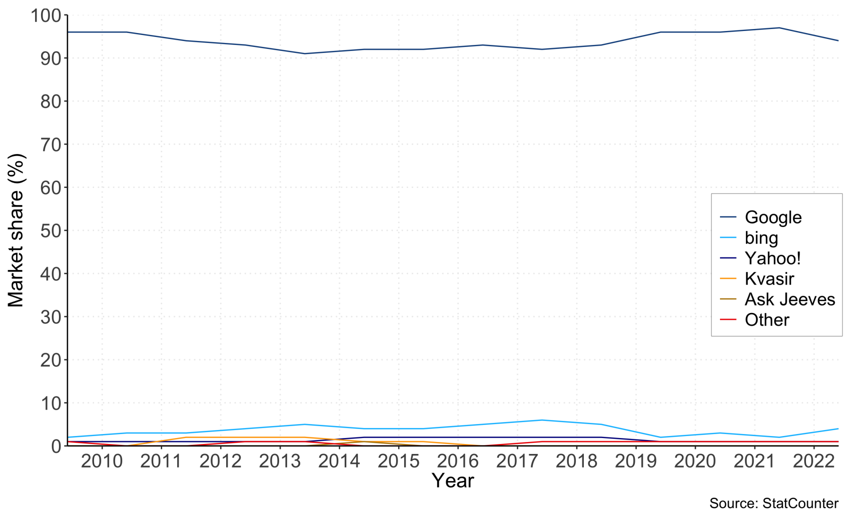
Figure 3.2: Number of internet users in Norway from 2000 to 2021



Search Engine Market Share

Google is Norway's most used search engine, with a 97% market share in 2021 (Statista, 2021). The market share of Google has been stable and dominant since 2009 (see Figure 3.3). Combined with the internet penetration, the search volume from Norway is large and stable for the Google search engine. As a result, we can assume this data represent public sentiment since 2010.

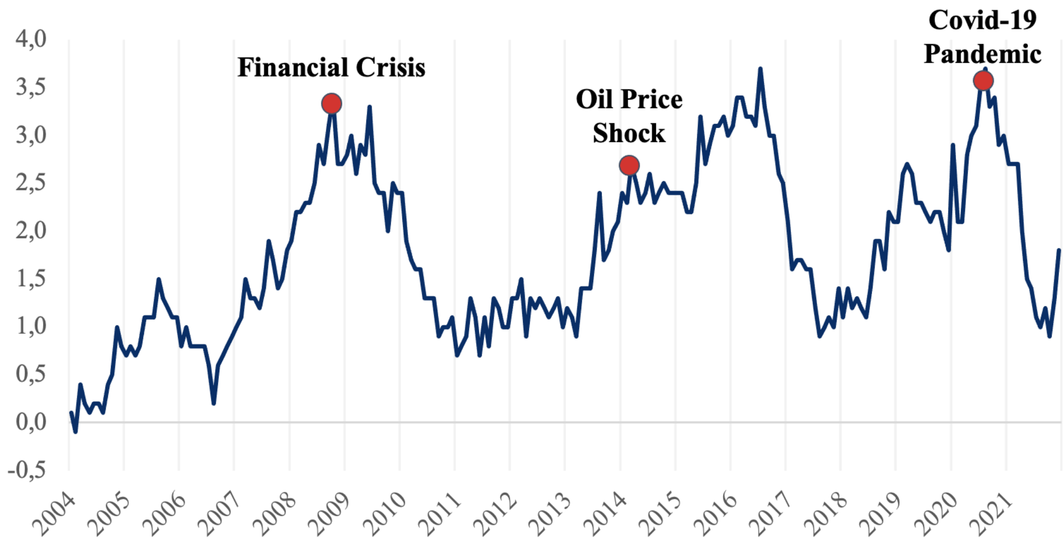
Figure 3.3: Market share of search engines in Norway from 2009 to 2022



Specific Events

When choosing the time series length, we want to include events that can provide useful information and increase our predictions' precision. For example, during the time in which Google Trends is available, three major economic crises have affected inflation in Norway (see Figure 3.4). These are the financial crises in 2008-2009, the oil price shock in 2014, and the Covid-19 pandemic in 2020-2021 (Lie, 2021; SSB, 2021; The Norwegian Central Bank, 2015). These are all of interest for our analysis.

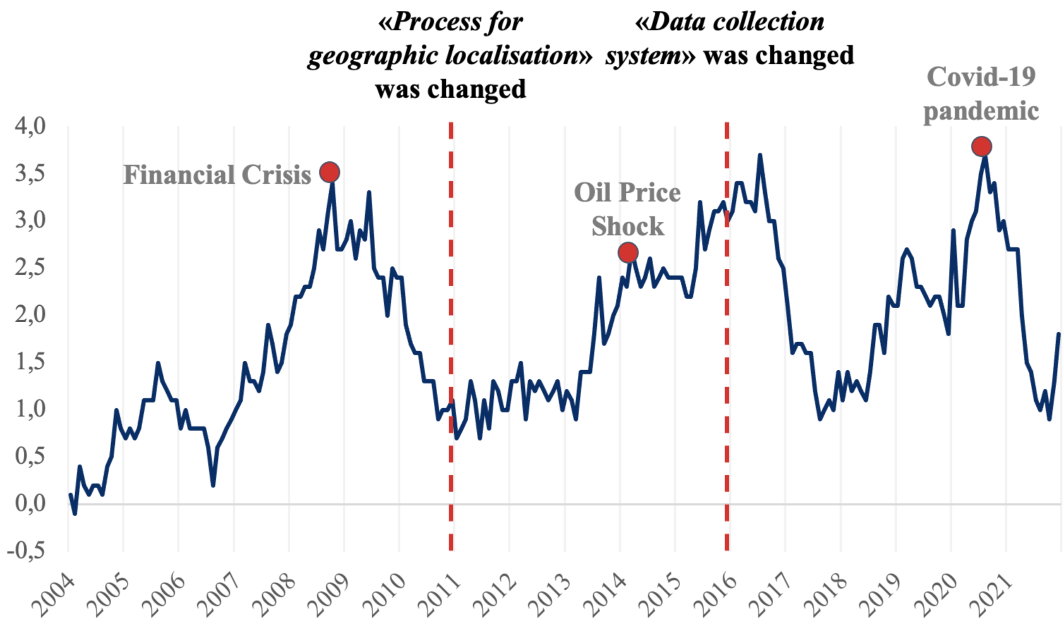
Figure 3.4: Significant events for CPI-ATE from 2004 to 2022



Structural Breaks

Google adjusted its data collection process, which induced breaks in January 2011 and January 2016 (see Figure 3.5). Both of these breaks are briefly documented on the website of Google Trends. In 2011, the *"process for geographic localisation"* was adjusted, while in 2016, the *"data collection system"* was *"improved"* according to them (Woloszko, 2020). The existing literature does not address these breaks as issues. However, Woloszko (2020) takes a different stance. The paper argues that these breaks can lead to the occurrence of outliers affecting the validity of the whole data set. As a result, we want to exclude these breaks.

Figure 3.5: Structural breaks in Google Trends data and significant events for CPI-ATE from 2004 to 2022



Number of Observations

When choosing data originating from search engines, it is essential to evaluate if the number of searches is large enough in absolute terms to create a representative sample (Hyndman & Athanasopoulos, 2018). The rule of thumb for forecasting with ARIMA models is that at least 50 observations should be included (Box & Tiao, 1975). However, it is preferable to have more than 100 observations. Therefore, we want to include more than 100 monthly observations, which implies a time series ranging back to at least 2015.

Chosen Period

For the reasons stated above, we downloaded Google search data from January 2011 to December 2021. As such, the Google data will include one structural break.

3.3.2 Frequency

Google search data are available at daily to yearly frequencies (see Section 3.2.1). Choosing a higher frequency would yield a better description of real-time micro behaviour. However, the data are not all systematically available at a higher frequency than the frequency we wish to predict. This unavailability complicates the use of higher frequency data than monthly. In addition, monthly frequencies are sufficient for the purpose of this thesis. We are more concerned with investigating if there is value in adding real-time micro behaviour search queries. This monthly data can close the ten-day gap of official inflation data to improve nowcast performance. As a result, we apply monthly frequencies.

3.3.3 Choice of Keyword

The purpose of Google search data is not for statistical analysis (Woloszko, 2020). This feature of the data calls for careful consideration when selecting keywords. Therefore, it is necessary to minimise the ambiguity and variation in the data by including keywords that we know with great certainty are directly linked with the phenomenon we are trying to predict. As a result, we have chosen keywords based on five criteria. These are listed and described in Table 3.3 below.

Table 3.3: Keyword selection method

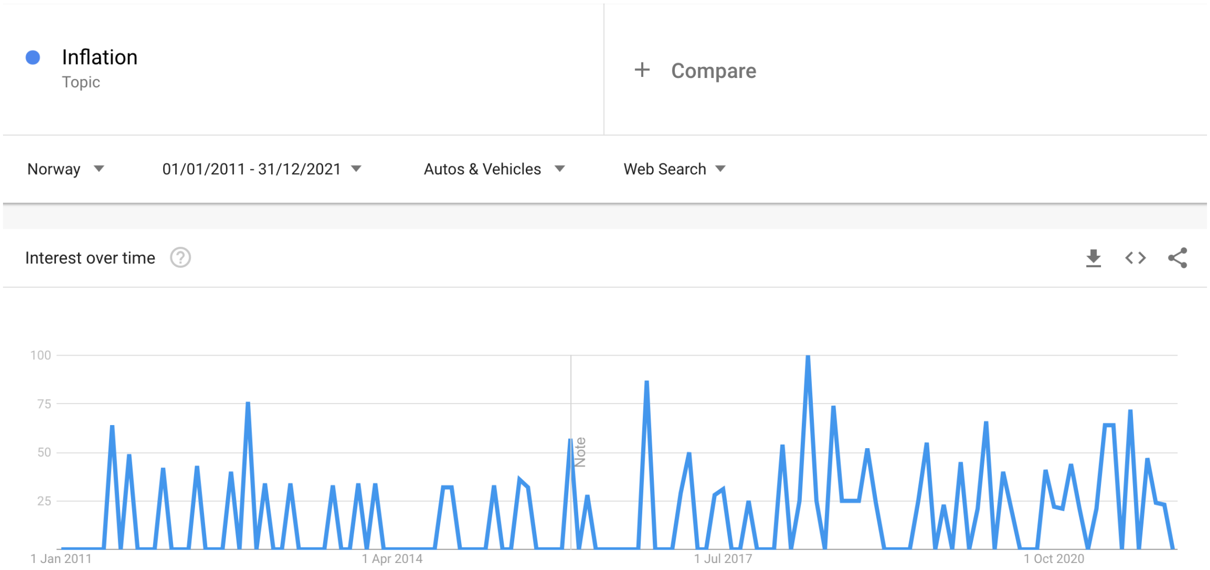
Criteria	Explanation
Empirical relevance	Only keywords that have, through empirical research, shown to be linked to inflation and to improve forecasts of inflation are considered.
Theoretical relevance	Downloaded keywords have a theoretical and empirical link to inflation to ensure validity in our analysis.
Quality of data	Keywords that have an unstable number of searches are eliminated. These searches contain zero values and large volatility over short periods on Google Trends.
Number of searches	>1000 average searches each month to ensure that our analysis has a sufficient number of observations. Evaluated by using Google Adwords.
Consistency	The meaning of the keywords should be stable across time to ensure it represents the same phenomenon for the whole period.

Only two keywords have been applied in existing research that has proved to be of consistent utility when nowcasting inflation utilising Google search data (Li, Shang, Shouyang, & Ma, 2015; Samantha, 2019). These keywords are *"price"* and *"inflation"*. In addition, multiple semantically linked extensions and versions of these two keywords have empirically shown to improve nowcasts of inflation. For instance, *"rising prices"* and *"increase in price"* are two examples. Therefore, we start our keyword selection by considering these two keywords.

Keyword "Inflation"

Variations of the keyword *"inflation"* have proved to increase the predictive ability of models (Samanta, 2019). In Norway, the number of searches for *"inflation"* are both small and volatile (see Figure 5.1). Filtering this search term for categories illustrates that the search term is unsuited for our purposes as it includes multiple periods with an insignificant number of searches.

Figure 3.6: Google search query for keyword Topic *"inflation"*, full sample period by category *"Autos & Vehicles"*



Keyword "Price"

Li, Shang, Shouyang, & Ma (2015) illustrate in their paper that keywords and search terms that contain the word *"price"* can be added to forecasting models to improve the precision of forecasts of inflation. Samanta (2019) illustrates the same result for nowcasting inflation in India, where she concludes that *"price"* has a solid ability to track inflation and is strongly correlated with inflation rates. Hence, *"price"* has been shown to have some utility in similar nowcasting models as the one we are creating. The economic intuition behind including *"price"* as a keyword is that people using this keyword in their queries may be interested in searching for information on the price of an item before purchasing the good.

Therefore, an increase in the relative number of searches for this term likely illustrates an increase in demand. Increasing demand that rises faster than supply increases inflation (Dubay, 2022).

The number of absolute monthly searches for *"price"* is higher than 1,000 (see Appendix A1.1). Simultaneously, the time series are stable and do not contain any zero values between 2011 and 2022 for any Google Category. None of these criteria prohibits *"price"* from being added to the nowcast model.

The keyword *"price"* circumvents the problem of changing meaning of keywords (Woloszko, 2020). The meaning of price is constant for the whole period, and the word's meaning will likely not change in the future. This increases the robustness of the model as it explains the same phenomenon across time.

However, *"price"* is a homonym as it is spelt the same way in Norwegian as the word *"award"*. This double-meaning may reduce the robustness of the model as searches for the *"award"* are likely not linked to inflation and will create noise in our data. Consequently, it is preferable to work with Google Categories rather than on specific terms (Goel, Hofman, Lahaie, Pennock, & Watts, 2010). Google Categories allow us to specify within which domains we want to analyse the search volume for the word *"price"*. This feature solves the issues with homonyms. In addition, it circumvents the issue of the choice of multiple keywords, which is subject to subjectivity in addition to being a manual task.

As a result of our analysis, we choose to include only the keyword *"price"* in our model. To compensate for the lack of keywords, we use the feature, Topic, to capture all related search terms. Using this feature means that all searches which contain the word *"price"* is included. For instance, *"rise in prices"*. At the same time, we gather time series of the word *"price"* for all divisions of inflation by using the category feature of Google Trends. The intention behind this is that we can analyse if Google Trends have a larger effect on some specific divisions of inflation.

3.3.4 Choice of Google Categories

In order to analyse more granular levels of inflation, we link each division of inflation (COICOP) to its corresponding Google Category (see Table 3.4). We analyse each division's weights and contents to ensure our Google categories represent the same goods and services as the inflation divisions.

Some of the corresponding categories consist of multiple sub-categories, which constitute roughly the same weight (see Appendix A2.1). For instance, division 2 - alcoholic beverages and tobacco, consist of both “alcoholic beverages” and “tobacco", where they respectively constitute 63% and 37%. Therefore, to make sure that the Google Indicator will capture most of the variation in the inflation in the divisions, we identified two Google Categories which we added to create the indicator. These two categories are: *"alcoholic beverages"* and *"tobacco products"*.

Table 3.4: Divisions of Inflation (COICOP) and their corresponding Google Category

	Google Categories	
Division	Primary	Secondary
01 Food and non-alcoholic beverages	Cooking & Recipes	Non-Alcoholic Beverages
02 Alcoholic beverages and tobacco	Alcoholic Beverages	Tobacco Products
03 Clothing and footwear	Apparel	
05 Housing, water, electricity, gas and other fuels	Real Estate	
05 Furnishings, household equipment and routine maintenance	Home & Garden	
06 Health	Health	
07 Transport	Autos & Vehicles	Transportation & Logistics
08 Communications	Internet & Telecom	
09 Recreation and culture	Hobbies & Leisure	
10 Education	Education	
11 Restaurants and hotels	Restaurants	Hotels & Accommodations
12 Miscellaneous goods and services	Beauty & Fitness	Social Services

The choice of categories is based on a comparative analysis of each inflation subgroup's components and the existing Google Categories (SSB, 2022; Trasborg, 2017). The components of the Google Categories are identified and compared with the components of the divisions of inflation. This comparison is performed to identify if the sub-levels of the Google Categories contain the same items as the divisions of inflation. To limit the number of Google Indicators added to the analysis and to prohibit overlapping categories, we only include Google Categories at the first or second hierarchical level of the classification (see Appendix A3.1). Another reason for only using categories from these hierarchical levels is to increase the number of searches captured by each category. This increases the number of observations. In addition, none of the Google Categories stemmed from the same primary category. This consideration ensures that none of the categories contains the same inflation component.

The use of Google Categories is beneficial as it allows us to predict and analyse a more granular level of inflation which can be used to understand what division of inflation has the most benefit from adding Google search data. In extension, it makes it possible to analyse and predict specific components of inflation.

3.3.5 Data Collected

We download seventeen time series for each listed Google Category in Table 3.4. The time series consists solely of the topic *"price"* from Google Trends. Further, the data are averaged in order to circumvent some of the implications from normalisation.

4 Models

This chapter explains the statistical models and methods applied in the thesis. The first section presents the autoregressive model used to investigate if Google search queries contain information that improves predictive ability. Subsequently, we present the model used for nowcasting inflation. This includes introducing the basic properties of an autoregressive integrated moving average model and how it predicts future values. Lastly, we describe the estimator of performance evaluation.

4.1 Autoregressive Models

An autoregressive process (AR) is a tool to predict future values of any variable of interest using past values of the same variable (Hyndman & Athanasopoulos, 2018). The term autoregressive specifies that the model regresses the dependent variable linearly on its lagged values. The lagged values are used as predictors, where the number of lags decides the order of the model. For example, a process with one lag is a first-order AR model, while a process with two lags is an AR model of second order.

The simplicity of the autoregressive process stems from the fact that lagged values alone explain changes in the dependent variable. However, the error term captures random signals with equal intensity at different frequencies, known as white noise. The model is simple but often performs adequately compared to more advanced models (Chan, 2011).

In general, an autoregressive process of order p is expressed as follows:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t \tag{4.1}$$

Where y_t is the dependent variable of interest, c is a constant, ϕ is the coefficient for each lagged value i, and ϵ_t is white noise.

One of the simplest yet insightful models in forecasting is the AR(1) process. Equation 4.2 summarises the autoregressive process described above, with a lag order of 1. For this model, the dependent variable is solely regressed on itself one period ago, the first lag.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \tag{4.2}$$

The first step to answering the research question of the thesis relies on this model.

4.2 ARIMA Models

An Autoregressive Integrated Moving Average (ARIMA) model uses historical variations in an individual time series to provide forecasts (Wooldridge, 2015). It comprises an autoregressive (AR) and a moving average (MA) element. The AR component, as described above, is a time series function of lagged values. The MA term, on the other hand, is a combination of random disturbances that captures the average change in the series over time (Hyndman & Athanasopoulos, 2018). It is one of the most widely used forecasting approaches for univariate time series.[5] The framework can be applied to different time series dynamics by combining AR and MA models with a factor for differencing, as stationarity is required. This requirement means that the properties of the time series do not change across time (Box & Jenkins, 1970).

Consider a general ARIMA for a given stochastic time series process y_t:

$$y_t^* = \Delta^d y_t \tag{4.3}$$

$$y_t^* = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{4.4}$$

where d is the number of differentiations required to achieve stationarity and c is an intercept term. p is the number of past values of y_t, and ϕ is the coefficient for these past values. In the last term of this model, ϵ_t is the contemporaneous error term assumed to be identically and independently distributed (IID), while q is the number of past error terms, and ψ is their related coefficient.

By choosing specific values for p and q, we arrive at the constituent models from the ARIMA framework. For instance, for a given d, when the $p > 0$ and $q = 0$, we arrive at

[5] A univariate time series to a time series that consists of single (scalar) observations recorded sequentially over equal time increments.

an autoregressive model where the time series is modelled by solely using past values. On the other hand, to arrive at a moving average model where the series is modelled by using values of the error term, the model has to be of the order $p = 0$ and $q > 0$. However, the combination of $p > 0$ and $q > 0$ gives us an ARIMA model. In this instance, the series is modelled by using both past values of itself and previous error terms.

4.2.1 ARIMAX

ARIMA processes with exogenous variables are a multivariate model denoted ARIMAX (Rachev, 2007).

This process is expressed as follows:

$$y_t^* = c + \sum_{i=1}^{r} \beta_i x_{i,t} + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{4.5}$$

with r different exogenous variables, $x_{1,t}$, $x_{2,t}$, ... , $x_{r,t}$ affecting y_t. β_i is the coefficient for the exogenous variables x_i.

4.2.2 SARIMAX

Seasonality violates ARIMA models' assumption of stationarity. This seasonality can be controlled for by using a seasonal ARIMA model or SARIMA. Below is the general equation for this model, including an exogenous variable:

$$y_t^* = c + \sum_{i=1}^{r} \beta_i x_{i,t} + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{i=1}^{P} \Phi_i y_{t-si} + \sum_{j=1}^{Q} \Theta_j \epsilon_{t-sj} \tag{4.6}$$

The seasonal model fits an additional set of autoregressive and moving average components on lags, which is offset by some number of lags s. This is the frequency of the seasonality.

4.3 Forecast Performance

To evaluate the performance of the forecast, we use the root-mean-square-error (RMSE). This metric squares the errors before averaging, putting a relatively high significance on large errors. Large errors are undesirable in forecasting (Bjørnland, 2015). As a result, RMSE is a popular and fitting metric for evaluating the performance of forecasts.

RMSE provides a score that indicates the fit of the forecast. If a score of zero is achieved, then the model is a perfect fit as the goal of the forecast is to minimise the RMSE. The estimator is expressed as follows:

$$\text{RMSE} = \sqrt{\frac{\sum(y-\hat{y})^2}{n}} \tag{4.7}$$

where y is the observed values, $\hat{y}$ is the predicted values of y, and n is the number of periods.

5 Forecasting Methodology

This chapter is dedicated to describing the forecasting methodology. Our point of departure is the simple autoregressive model of order one. First, we perform a preliminary investigation by adding the Google Indicators to each AR(1)-model to identify if search queries represented by Google contain valuable information that improves the predictive power of the baseline model. The analysis is then extended to identify the most accurate model for predicting inflation, with and without Google Indicators. This identification is conducted within the acknowledged Box-Jenkins methodology. Lastly, the predictive ability of the most accurate models is subject to assessment in a one-month ahead out-of-sample nowcast.

5.1 AR (1)-Test of Content

The preliminary analysis consists of twelve AR(1) processes as baseline models, representing each division of CPI (see Table 3.4). Each one of these AR(1) processes is then supplemented with representative Google Indicators in order to analyse if the prediction error (RMSE) is reduced (Varian & Choi, 2009).

We do this for two reasons. First, the existing literature on Google search queries performs this preliminary test to check if the search queries represented by Google contain any information of interest (see Section 2.2). Second, the existing literature on inflation modelling suggests that AR(1) models will likely outperform any known model when forecasting inflation (Chan, 2010). Therefore, if the RMSE is reduced, then Google Trends data contain valuable information for predicting inflation.

The AR(1) models are expressed as follows:

$$\text{Baseline AR(1)} = \pi_t^b = c + \phi_i y_{t-1} + \epsilon_t \tag{5.1}$$

$$\text{Extended AR(1) (single)} = \pi_t^s = c + \phi_i y_{t-1} + \omega_i GI_{it} + \epsilon_t \tag{5.2}$$

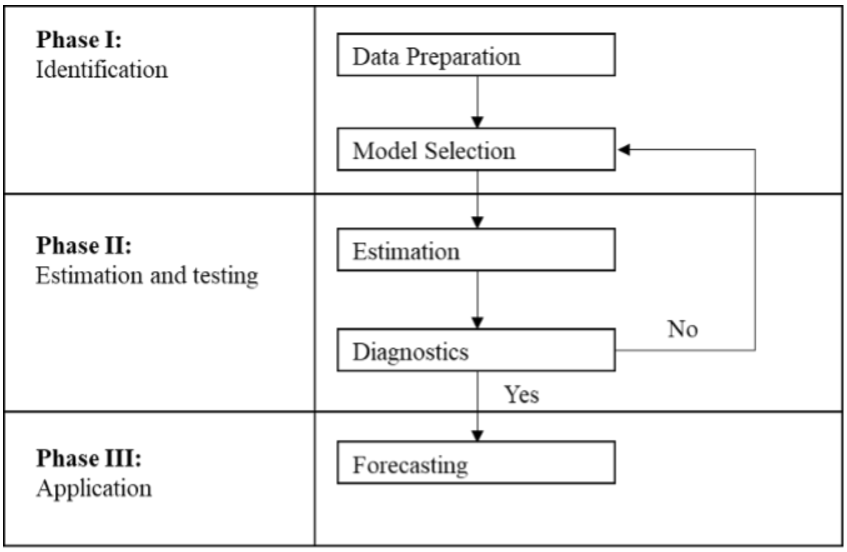
$$\text{Extended AR(1) (dual)} = \pi_t^d = c + \phi_i y_{t-1} + \omega_i GI_{it} + \omega_j GI_{jt} + \epsilon_t \tag{5.3}$$

where π_t^z is the inflation rate for model z at time t. c is a constant, ϕ_1 is the coefficient for the lagged value of order 1, while ϵ_t is the error term at time t. Further, ω_{it} and ω_{jt} are the coefficients for the Google Indicators, GI_{it} and GI_{jt}, at time t.

5.2 The Box-Jenkins Method for Model Selection

The Box-Jenkins method applies ARMA and ARIMA models to forecast time series based on lagged values (Hyndman & Athanasopoulos, 2018). The framework separates into three phases (see Figure 5.1).

Figure 5.1: Box-Jenkins methodology for model selection



The identification phase commences by collecting and examining the relevant data. Initially, the data are inspected graphically. The inspection is then extended to a statistical approach that aims to decompose the statistical properties of the data. Irregularities such as non-stationarity will require treatment.

Once the stochastic process of stationarity is ensured, the data is subject to model construction. Alternative estimators for out-of-sample prediction error will be supplemented to assert an appropriate and unbiased identification process.

The framework then proceeds to examine the diagnostics of the identified models. This thesis will focus on residual tests to ensure random distribution in the data. Once the data is independently distributed, the thesis will focus on out-of-sample forecasts. In the Box-Jenkins framework, this forecasting process is considered circular rather than sequential. The fourth and fifth steps feedback to the third to ensure the robustness of the selected models. The final procedure is a pseudo out-of-sample comparison between the predicted values and an omitted validation set. This comparison is made to evaluate the predictive ability of the models constructed in-sample.

5.2.1 Phase I: Identification

The introductory phase comprises two steps; data preparation and selection of the appropriate models. The former aims to identify and deal with non-stationarity in the dependent variable to determine the necessity of differentiating *(d)* the time series. The latter involves analysing the ACF and PACF plots to determine the lag order of the autoregressive component *(p)* and the order of the moving average component *(q)*. This phase summarises the complete ARIMA*(p,d,q)* process.

Data Preparation

In line with the Box-Jenkins methodology, the time series must be stationary for the forecast to be valid. The data will be subject to differentiating if the time series violates stationarity as it is impossible to draw inference on the model's predictive ability before this is achieved (Wooldridge, 2015). Thus, providing valid and unbiased predictions becomes contingent on the stationarity of the data.

Stationarity infers that the statistical properties of the time series do not change with time. A time series is considered stationary if it meets three defined conditions (Wooldridge, 2015). First, the mean of the time series is constant over time. Second, the variance is constant over time. Third, the autocovariance must be constant over time. The third condition infers that the time series exhibits no trend component.

These conditions are expressed as follows:

1. Constant mean (μ) for all t: $E(y_t) = \mu$
2. Constant variance (σ) for all t: $Var(y_t) = \sigma^2$
3. Constant autocovariance: $Cov(y_t, y_{t-s}) = \mu$

Overlooking stationarity has severe implications for the forecasting procedure. A nonstationary time series allows only for studying one period individually, as each period of the time series will be for a particular episode. Thus, unless differentiated, it could lead to spurious results. This spuriosity makes nonstationary time series unreliable and of negligible practical value for forecasting purposes.

One approach to formally detect the presence of nonstationary is an augmented Dickey-Fuller test (ADF). Specifically, the ADF test investigates if the time series in question exhibits the presence of a unit root to determine whether the time series is stationary or nonstationary (Wooldridge, 2015). Thus, the ADF test aims to identify the *(d)* and *(D)* in the general ARIMA*(p,d,q)x(P,D,Q)*.

Model Selection

Once stationarity is dealt with, and the degree of differentiation is identified, the Box-Jenkins framework aims to identify the appropriate AR*(p)* and MA*(q)* components. This selection is conducted through a thorough analysis of the partial autocorrelation function (PACF) and the autocorrelation function (ACF). However, there is no formal test for identifying the lag order *(p)* or the order of moving average *(q)*. Identification develops into a subjective exercise when p and q simultaneously are non-zero. Thus the experience and training of the forecaster become increasingly relevant. Several models can be unveiled as plausible alternatives for the time series in situations like these. However, to ensure a better statistical fit, using the least number of parameters in an ARIMA model is recommended to achieve a sufficient representation (Box, Gwilym, & Reinsel, 2008).

Further, general practices urge the forecaster to issue carefulness when selecting ARIMA models. For example, one common pitfall is that the forecaster over-specifies the model in what is known as data mining. Data mining implies pattern recognitions and improved explanatory power based on in-sample selection criteria. One such pitfall can be model selection based on RMSE as the single criteria, which has the potential to produce questionable out-of-sample predictions. Thus, any in-sample selection criteria exhibit a bias that is not likely to hold for the out-of-sample prediction.

To circumvent biased in-sample estimators, the selection procedure is extended to incorporate estimators of prediction errors that penalise in-sample residual variance. Estimators with such capabilities consider the degrees of freedom for the model in question.

In line with the broad literature on the field, this thesis rests on two commonly applied criteria, namely the Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC).

The AIC and BIC have the same objective and share several attributes. Ultimately, these estimators of prediction error aim to minimise the in-sample residual sum of squares. Moreover, both estimators carry a penalty term for the total number of parameters estimated. The model of better fit based on the two criteria is the one that yields the lowest estimator of prediction error.

The Akaike's Information Criterion (AIC) is expressed as follows:

$$\mathrm{AIC} = \log\left(\frac{\sum \hat{\epsilon_i}^2}{N}\right) + \frac{2k}{N} \tag{5.4}$$

Whilst the Bayesian Information Criterion (BIC) is expressed as follows:

$$\mathrm{BIC} = \log\left(\frac{\sum \hat{\epsilon_i}^2}{N}\right) + \frac{k\log(N)}{N} \tag{5.5}$$

Where $\sum \hat{\epsilon_i}^2$ is the residual sum of squares, N is the number of observations, and k represents the number of coefficients estimated.

It is possible to draw some conclusions about the distinctions between the AIC and BIC based on Equations 5.4 and 5.5. For observations (N) larger than 7; the BIC estimator tends to penalise models more severely than the AIC. Mathematically, this occurs when $\log N > 2$. For this reason, the BIC is more meticulous than the AIC when selecting the models. However, applying both estimators when identifying ARIMA models is recommended, as they, in some cases, are complements rather than substitutes.

For our purposes, as explained above, the estimators ensure more robust models when performing out-of-sample predictions. However, both estimators have some inconsistencies. First, as there are no guidelines for testing, comparison between models is strictly based on the estimated values alone. Second, comparing these estimators is recommended for similar ARIMA models only. Lastly, output values tend to be of marginal difference in several cases.

The absence of statistical tests to compare the two estimators makes the comparison somewhat subjective. On the one hand, Makridakis et al. (1998) favour the BIC due to the overparameterisation of the AIC.[6] On the other hand, Bjørnland (2015) argue that the AIC is preferred for choosing forecasting models, while BIC is preferred for drawing inference. Thus, there is some ambiguity in the selection criteria. Throughout this thesis, we limit the scope of the ARIMA models to $p = 4$ and $q = 4$, as the models can become too extensive.

Auto.arima

This thesis processes a large amount of data. To limit the consequence of human error, we support our model selection procedure with the auto.arima function R package *forecast*.[7] The function provides an order of identification for each series. It is beyond the scope of this thesis to discuss the detail of the automatic model identification procedure. Instead, we refer to the literature by Hyndman & Khandakar (2008) for further details of the procedure.

5.2.2 Phase II: Estimation and Testing

Estimation

Once a provisional model has been identified, we apply statistical software to generate estimates for the parameters, including a test for significance for each parameter in the model (Linden, Adams, & Roberts, 2003). If any of the parameters are insignificant, we eliminate them to improve the model's fit.

Diagnostics

The final step in Phase II is a diagnosis to ensure that the selected models fulfil the requirements for a univariate time series. The time series are subject to a test of whether the residuals are white noise. If the selected models are an appropriate fit, the ACF plot should show no significant autocorrelations among the residuals. However, if there are significant autocorrelations, it will require the re-identification of models.

In this thesis, the preferred assessment tool is a formal portmanteau test known as the Ljung-Box Q-test.[8] The Ljung-Box test is a statistical test that confers whether any

[6] Overparameterised in the sense of too many AR and MA terms in the specified ARIMA model.
[7] Arguments in the auto.arima()-function are set to default.
[8] In statistical hypothesis testing, A portmanteau test is a type of statistical hypothesis test in which

group of autocorrelations of a time series are different from zero (Box & Pierce, 1970). In specific, it is a test of the following hypothesis that assess any linear dependencies:

H0: The time series is either random, white noise or i.i.d

HA: The time series is non-random (serial correlation)

The hypothesis test is based on the autocorrelation plot of the time series. For example, in a white noise series, 95 per cent of the spikes in the ACF will be observed within the critical values.

Statistically, the hypothesis is tested with the test statistic Q in Equation 5.6:

$$\mathrm{Q} = n(n+2)\sum_{k=1}^{h}\left(\frac{r_k^2}{n-k}\right) \tag{5.6}$$

Where n is the number of observations, h is the number of lags considered, and r_k is the accumulated sample autocorrelations of the residuals.

If the test null hypothesis is true, then none of the autocorrelation coefficients proves to differ from zero statistically. This holds for all lags up until h. Thus, the data are random, and the residuals are white noise. On the opposite, the null hypothesis is rejected, and the model shows a lack of fit if $\mathrm{Q} > \chi^2_{1-\alpha,h-m}$, where $\chi^2_{1-\alpha,h-m}$ is the value found in the chi-square distribution for significance level α and h-m degrees of freedom where m is the number of parameters in the model.

A Q-statistics within the extreme 5% of the right-hand tail is considered non-white noise (Hyndman R. J., 2014). As for the number of lags to include, there is no standard practice. However, different approaches have been recommended, whereby one argues 15 lags (Makridakis, 1998), while others argue 20 to 24 (Pindyck & Rubinfeld, 1998). For the purposes of the thesis, this distinction seems to be of marginal difference, with an identical outcome.

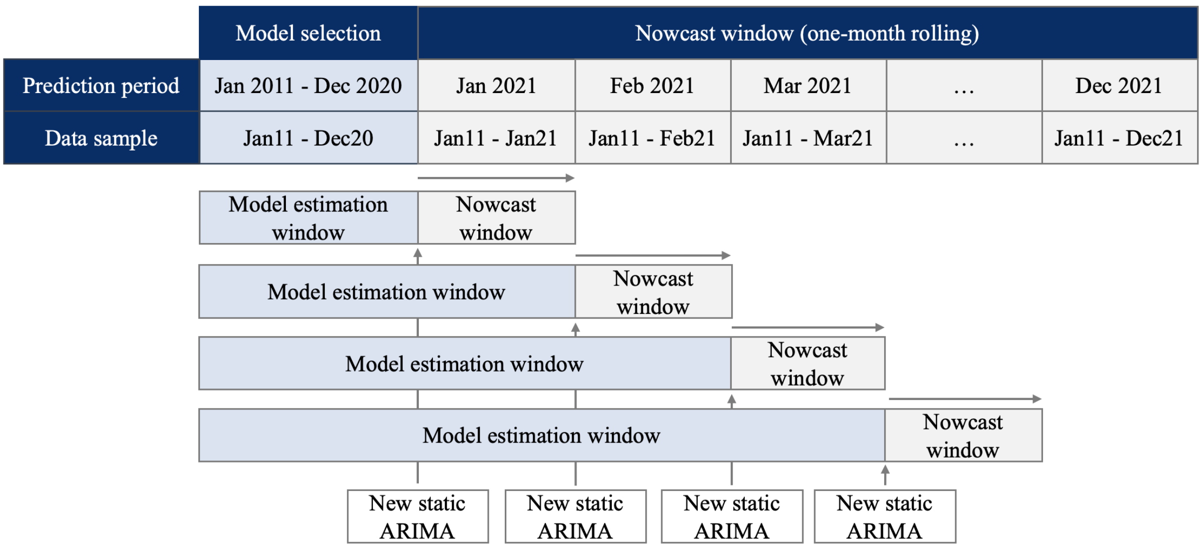
the null hypothesis is well specified, but the alternative hypothesis is more loosely specified.

5.2.3 Phase III: Application

Nowcasting

Once the foundation is laid, it is time to generate the out-of-sample nowcasts. The procedure is to predict divisions of inflation h-steps ahead, where h denotes the number of periods we are going to forecast. Moreover, our approach is static, which means that all models are subject to re-estimation once new data is available (Stock & Watson, 2008). As such, we ensure that all available information is reflected in the model at any given time (see Figure 5.2). Finally, this process is repeated until the end of the out-of-sample period. Thus, for each division of inflation, the exercise will yield twelve $h = 1$ predictions.

Figure 5.2: The figure shows the pseudo out-of-sample monthly nowcast method.



The in-sample period is used to estimate the extended ARIMA models with Google Indicators and the benchmark non-extended AR(1). This sample starts in January 2011 and ends December 2020. For the out-of-sample period, the window ranges from the first month of 2021 up to and including December 2021. We have chosen this window to constantly have the latest data included in our model to increase the number of observations and to use the newest information, increasing our nowcasts' precision.

Robustness of Models

To further investigate if Google search data can outperform the AR(1), we compare extended models to their respective baselines across two dimensions. First, we compare the models in terms of prediction error. The RMSE measures this metric (see Section 4.3). The prediction error of the extended model is measured by the deviation between the predicted values and CPI ex-post.

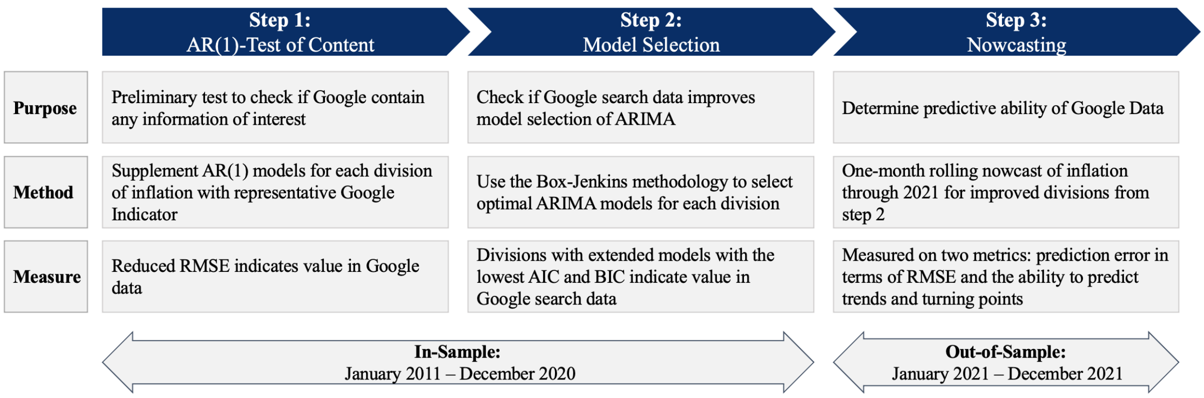
Second, we compare the extended and baseline models' ability to identify trends and turning points. This comparison aims to evaluate if Google search queries improve the ability to identify short-term changes in inflation. Specifically, the models are compared in terms of the number of months they can predict the trends and turning points of CPI ex-post.

In conclusion, these two dimensions will provide a framework for assessing the predictive ability of ARIMA models extended with Google search data.

Three-Step Approach of Analysis

This thesis is designed to perform pseudo-out-of-sample predictions. In particular, we perform static one-month ahead rolling predictions out-of-sample. These nowcast results are the benchmark against a validation set of recorded CPI. To arrive at this stage, we follow a three-step approach (see Figure 5.3):

Figure 5.3: Three-step process for out-of-sample nowcast



First, we conduct a preliminary test of content by constructing a set of AR(1) models for each division of inflation. These represent the AR(1) baselines. We then extend our baseline models by adding Google Indicators to investigate if Google search queries are information value-adding. We aim to identify if these extended models reduce the RMSE compared to their respective baselines.

The next step extends the analysis by using the Box-Jenkins methodology described in this chapter to identify ARIMA models for each division of inflation. Models are ranked and selected by the estimator of prediction error they yield through the in-sample period from January 2011 to December 2020.

Based on Step 2, the final step proceeds with the divisions where extended models were identified as a better fit for prediction than their respective baselines. Subsequently, we perform an out-of-sample one-month rolling nowcast through 2021 to determine the predictive ability of Google data. In addition, we compare our extended ARIMA models to AR(1) models before benchmarking both against CPI ex-post.

6 Results

In this section, we present the results from the three steps of the thesis (see Figure 5.3). In the first step, we conduct a preliminary test to investigate the information value-added of the Google search data. Step 2 proceeds by using the acknowledge Box-Jenkins methodology to identify ARIMA models for each division of inflation. Lastly, based on Step 2, we perform an out-of-sample one-month rolling nowcast through 2021 to determine the predictive ability of Google data. In addition, we compare our extended ARIMA models to AR(1) models for the same divisions of CPI before benchmarking both against CPI ex-post.

6.1 Step 1: In-Sample AR (1)-Test of Content

In this section, we add the selected Google Indicators to AR(1) processes of the baselines to construct the extended models. Further, we compare the root-mean-square error (RMSE) of the extended model to the baseline model to investigate if Google Indicators can reduce standard deviations of residuals.

Table 6.1: RMSE and change in RMSE of AR(1) models after adding Google Indicators

	RMSE						
Division	Baseline	Extended 1	*% change*	Extended 2	% change	Extended 3	*% change*
COICOP 01	1.15069	1.15060	-0.008%	1.13960	-0.964%	1.13816	-1.089%
COICOP 02	0.38230	0.38194	-0.092%	0.37771	-1.200%	0.37670	-1.464%
COICOP 03	1.91881	1.91033	-0.442%				
COICOP 04	0.15882	0.15278	-3.801%				
COICOP 05	1.23015	1.22253	-0.620%				
COICOP 06	0.50631	0.50622	-0.018%				
COICOP 07	1.17643	1.16168	-1.254%	1.16087	-1.323%	1.15937	-1.450%
COICOP 08	1.09531	1.09248	-0.259%				
COICOP 09	0.48526	0.48469	-0.118%				
COICOP 10	0.69617	0.69108	-0.730%				
COICOP 11	0.39572	0.39529	-0.109%	0.39572	0.000%	0.39524	-0.122%
COICOP 12	0.30281	0.30265	-0.052%	0.30279	-0.006%	0.30264	-0.056%

The results from the preliminary test of content are illustrated in Table 6.1. It is evident that there is some information of interest in the Google search queries as all but one extended model reduces the RMSE of the baselines. In particular, COICOP 01, 02, 04 and 07 are notably improved by adding Google Indicators, while COICOP 06, 08, 09, 11 and 12 experience only minor improvements in terms of RMSE. In general, the results indicate valuable information in adding Google search data.

6.2 Step 2: In-Sample Model Selection

The results from the in-sample predictions between January 2011 and December 2020 are presented in Table 6.2. In total, the performances of extended models compared to their respective baselines are mixed. Extended models yield lower estimators of prediction error for six of the twelve CPI divisions. The AIC estimator ranks extended models as a better fit in four cases, whereas the BIC estimator draws the same conclusion for five of the six improved divisions. We find this result interesting as the BIC estimator by construction penalises multivariate models more than the AIC. For the scope of this paper, this has no implications.

Further, the estimators of prediction error are unanimous in ranking extended models as preferred in three of the six improved cases. These are the divisions measuring prices of food and non-alcoholic beverages (COICOP 01), clothing and footwear (COICOP 03), and housing, water, electricity, gas and other fuels (COICOP 04). We observe no improvement for the five divisions containing dual-category models.

Table 6.2: Model ranking ranked by (AIC) and BIC estimators

COICOP 01 - Food and non-alcoholic beverages							
AIC				BIC			
Rank	ARIMA	Model	Estimator	Rank	ARIMA	Model	Estimator
1.	(0,1,1)x(0,0,1)	Extended 1	335.67	1.	(0,1,1)x(0,0,1)	Extended 1	346.79
2.	(0,1,1)x(0,0,1)	Baseline	338.93	2.	(0,1,1)x(0,0,1)	Baseline	347.27
3.	(1,0,1)x(0,0,1)	Extended dual	338.98	3.	(0,1,1)x(0,0,1)	Extended 2	351.11
4.	(0,1,1)x(0,0,1)	Extended 2	339.99	4.	(1,0,1)x(0,0,1)	Extended dual	358.49
5.	(1,0,0)x(0,0,0)	AR(1)	380.97	5.	(1,0,0)x(0,0,0)	AR(1)	389.33

COICOP 02 - Alcoholic beverages and tobacco							
AIC				BIC			
Rank	ARIMA	Model	Estimator	Rank	ARIMA	Model	Estimator
1.	(2,1,2)x(0,0,1)	Baseline	76.85	1.	(2,1,2)x(0,0,1)	Baseline	93.53
2.	(2,1,2)x(0,0,1)	Extended 2	77.90	2.	(2,1,2)x(0,0,1)	Extended 2	97.35
3.	(1,0,0)x(0,0,1)	Extended dual	86.75	3.	(1,0,1)x(0,0,1)	Extended 1	101.13
4.	(1,0,1)x(0,0,1)	Extended 1	87.19	4.	(1,0,0)x(0,0,1)	Extended dual	103.47
5.	(1,0,0)x(0,0,0)	AR(1)	117.32	5.	(1,0,0)x(0,0,0)	AR(1)	125.69

COICOP 03 - Clothing and footwear							
AIC				BIC			
Rank	ARIMA	Model	Estimator	Rank	ARIMA	Model	Estimator
1.	(4,1,2)x(0,0,1)	Extended 1	447.55	1.	(4,1,2)x(0,0,1)	Extended 1	472.56
2.	(3,0,1)x(2,0,0)	Baseline	474.34	2.	(3,0,1)x(2,0,0)	Baseline	493.86
3.	(1,0,0)x(0,0,0)	AR(1)	503.82	3.	(1,0,0)x(0,0,0)	AR(1)	512.18

COICOP 04 - Housing, water, electricity, gas, and other fuels							
AIC				BIC			
Rank	ARIMA	Model	Estimator	Rank	ARIMA	Model	Estimator
1.	(1,1,0)x(0,0,1)	Extended 1	-115.23	1.	(1,1,0)x(0,0,1)	Extended 1	-101.33
2.	(1,1,0)x(0,0,1)	Baseline	-109.14	2.	(1,1,0)x(0,0,1)	Baseline	-98.03
3.	(1,0,0)x(0,0,0)	AR(1)	-92.44	3.	(1,0,0)x(0,0,0)	AR(1)	-84.08

COICOP 05 - Furnishings, household, and routine maintenance							
AIC				BIC			
Rank	ARIMA	Model	Estimator	Rank	ARIMA	Model	Estimator
1.	(0,1,2)x(1,0,1)	Baseline	340.19	1.	(0,1,2)x(0,0,1)	Extended 1	354.64
2.	(0,1,2)x(0,0,1)	Extended 1	340.74	2.	(0,1,2)x(1,0,1)	Baseline	356.86
3.	(1,0,0)x(0,0,0)	AR(1)	398.10	3.	(1,0,0)x(0,0,0)	AR(1)	406.46

COICOP 06 - Health							
AIC				BIC			
Rank	ARIMA	Model	Estimator	Rank	ARIMA	Model	Estimator
1.	(1,0,3)x(0,0,1)	Baseline	137.89	1.	(1,0,1)x(0,0,1)	Extended 1	156.64
2.	(1,0,1)x(0,0,2)	Extended 1	139.92	2.	(1,0,3)x(0,0,1)	Baseline	157.41
3.	(1,0,0)x(0,0,0)	AR(1)	184.27	3.	(1,0,0)x(0,0,0)	AR(1)	192.63

COICOP 07 - Transport							
AIC				BIC			
Rank	ARIMA	Model	Estimator	Rank	ARIMA	Model	Estimator
1.	(1,0,1)x(0,0,1)	Extended 1	328.63	1.	(0,1,1)x(0,0,1)	Baseline	344.77
2.	(2,0,0)x(2,0,0)	Extended 2	331.45	2.	(1,0,1)x(0,0,1)	Extended 1	345.36
3.	(2,0,1)x(2,0,0)	Extended dual	332.02	3.	(2,0,0)x(2,0,0)	Extended 2	350.97
4.	(0,1,1)x(0,0,1)	Baseline	336.43	4.	(2,0,1)x(2,0,0)	Extended dual	357.10
5.	(1,0,0)x(0,0,0)	AR(1)	385.83	5.	(1,0,0)x(0,0,0)	AR(1)	394.19

COICOP 08 - Communications							
AIC				BIC			
Rank	ARIMA	Model	Estimator	Rank	ARIMA	Model	Estimator
1.	(0,1,0)x(0,0,2)	Baseline	307.89	1.	(0,1,0)x(0,0,2)	Baseline	319.01
2.	(0,1,0)x(2,0,1)	Extended 1	309.02	2.	(0,1,0)x(2,0,1)	Extended 1	325.69
3.	(1,0,0)x(0,0,0)	AR(1)	370.63	3.	(1,0,0)x(0,0,0)	AR(1)	378.99

COICOP 09 - Recreation and culture							
AIC				BIC			
Rank	ARIMA	Model	Estimator	Rank	ARIMA	Model	Estimator
1.	(0,0,1)x(0,01)	Baseline	138.71	1.	(0,0,1)x(0,01)	Baseline	147.04
2.	(0,0,1)x(0,01)	Extended 1	140.69	2.	(0,0,1)x(0,01)	Extended 1	151.80
3.	(1,0,0)x(0,0,0)	AR(1)	175.55	3.	(1,0,0)x(0,0,0)	AR(1)	183.91

COICOP 10 - Education							
AIC				BIC			
Rank	ARIMA	Model	Estimator	Rank	ARIMA	Model	Estimator
1.	(1,0,0)x(0,0,0)	AR(1)	261.70	1.	(1,0,0)x(0,0,0)	AR(1)	270.06
2.	(1,0,0)x(0,0,0)	Baseline	261.70	2.	(1,0,0)x(0,0,0)	Baseline	270.06
3.	(1,0,0)x(0,0,0)	Extended 1	262.74	3.	(1,0,0)x(0,0,0)	Extended 1	273.89

COICOP 11 - Restaurants and hotels							
AIC				BIC			
Rank	ARIMA	Model	Estimator	Rank	ARIMA	Model	Estimator
1.	(2,0,0)x(0,0,1)	Baseline	77.53	1.	(2,0,0)x(0,0,1)	Baseline	91.47
2.	(2,0,0)x(0,0,1)	Extended 2	77.26	2.	(2,0,0)x(0,0,1)	Extended 2	93.98
3.	(2,0,0)x(0,0,1)	Extended dual	79.19	3.	(2,0,0)x(0,0,1)	Extended 1	96.17
4.	(2,0,0)x(0,0,1)	Extended 1	79.44	4.	(2,0,0)x(0,0,1)	Extended dual	98.70
5.	(1,0,0)x(0,0,0)	AR(1)	124.66	5.	(1,0,0)x(0,0,0)	AR(1)	133.03

COICOP 12 - Miscellaneous goods and services							
AIC				BIC			
Rank	ARIMA	Model	Estimator	Rank	ARIMA	Model	Estimator
1.	(0,1,1)x(2,0,0)	Baseline	26.88	1.	(0,1,1)x(2,0,0)	Baseline	38.00
2.	(0,1,1)x(2,0,0)	Extended 2	27.77	2.	(0,1,1)x(2,0,0)	Extended 2	41.66
3.	(1,0,1)x(2,0,0)	Extended 1	32.85	3.	(1,0,1)x(2,0,0)	Extended 1	52.36
4.	(1,0,1)x(2,0,0)	Extended dual	33.73	4.	(1,0,1)x(2,0,0)	Extended dual	56.03
5.	(1,0,0)x(0,0,0)	AR(1)	61.86	5.	(1,0,0)x(0,0,0)	AR(1)	70.22

6.2.1 Notable Performances

Table 6.3 clarifies the percentage improvement for each of the improved extended divisions, relative to baselines, for both estimators of prediction error. Two divisions stand out with the most significant improvement, namely clothing and footwear (COICOP 03) and housing, water, electricity, gas and other fuels (COICOP 04). Both divisions replicate the considerable AIC improvements when using BIC as the estimator of prediction error.

Further, we note that the extended model for the transport division (COICOP 07) demonstrates a relatively great improvement. All extended models for this division outperform the baseline model for the AIC estimator (see Table 6). While the extended models fail to replicate this for the BIC estimator, the baseline proves to be no more than marginally better than the best-extended model. This difference is of little importance regarding which model is the best fit between the extended and baseline models.

The remaining three improved divisions, namely Food and non-alcoholic beverages (COICOP01), Furnishings, household equipment and routine maintenance (COICOP 05), and Health (COICOP 06), all are improved relative to their baselines. However, the reduction in AIC and BIC is minor, and we cannot conclude that including Google search data improves model selection for these divisions.

Table 6.3: Improvement in prediction error for top-performing extended ARIMA model compared to baseline

	Relative to Baseline	
Division	AIC	BIC
01 Food and non-alcoholic beverages	-1.0%	-0.1%
02 Alcoholic beverages tobacco	1.4%	4.1%
03 Clothing and footwear	-5.6%	-4.3%
04 Housing, water electricity, gas and other fuels	-5.6%	-3.4%
05 Furnishings, household equipment and routine maintenance	0.2%	-0.6%
06 Health	1.5%	-0.5%
07 Transport	-2.3%	0.2%
08 Communications	0.4%	2.1%
09 Recreation and hotels	1.4%	3.2%
10 Education	0.4%	1.4%
11 Restaurants and hotels	0.3%	2.7%
12 Miscellaneous goods and services	3.3%	9.6%

6.2.2 Summary

Six of twelve divisions improved in terms of lower prediction error by adding Google Indicators. These are COICOP 01, 03, 04, 05, 06 and 07. These extended models outperformed their respective baselines for the in-sample period and are improved by including Google Indicators compared to the remaining six divisions. We use the results from the in-sample estimation as a selection criterion for the preceding step. As a result, we will nowcast inflation for each improved extended model out-of-sample.

6.3 Step 3: Out-of-Sample Nowcast

This thesis's scope is limited to nowcasting the six improved divisions of CPI, based on the results from Step 2. We perform nowcasts for extended ARIMA models and AR(1) models. Each division is assessed on nowcasting performance in terms of prediction error measured by RMSE. In addition, we assess each model's ability to predict the short-term changes in inflation in each period. This assessment is done by analysing the models' abilities to identify trends and turning points.

6.3.1 Performance: Prediction Error

We perform out-of-sample predictions for extended ARIMA and AR(1) models. The predicted values of the models are benchmarked against a validation set (CPI ex-post). Thus, the reported estimates of prediction error (RMSE) measure the real deviation. For four of six divisions, the extended model outperforms AR(1) in terms of RMSE (see Table 6.4). The extended models for COICOP 01, 04, 05, and 07 yields an average RMSE of 89.69, 15.93, 135.67, and 114.39 basis points, respectively. This represents a 4.70%, 26.30%, 15.27% and 9.43% improvement.

Table 6.4: RMSE nowcasted divisions of inflation

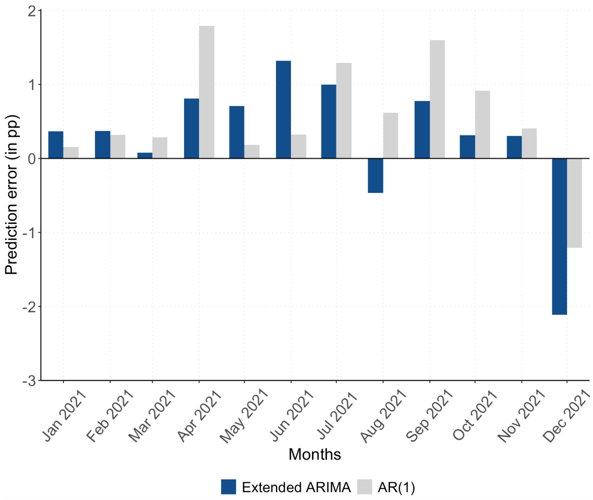
	RMSE		
Division	Extended ARIMA	AR(1)	Relative change
01 Food and non-alcoholic beverages	86.69	94.11	4.70%
03 Clothing and footwear	190.61	156.71	-21.64%
04 Housing, water, electricity, gas and other fuels	15.93	21.61	26.30%
05 Furnishing, household equipment and routine maintenance	135.67	160.12	15.27%
06 Health	294.08	278.95	-5.43%
07 Transport	114.39	126.31	9.43%

When comparing the monthly prediction error of the extended ARIMA and the AR(1) to CPI ex-post, we see that none of the models performs systematically better every month (see Figure 6.1). However, COICOP 01, 04, 05, and 07 have a lower deviation from actual inflation compared to AR(1) for no less than eight months. On average, the extended models have a mean deviation of 115 basis points from actual inflation compared to 118 basis points for the AR(1) models.

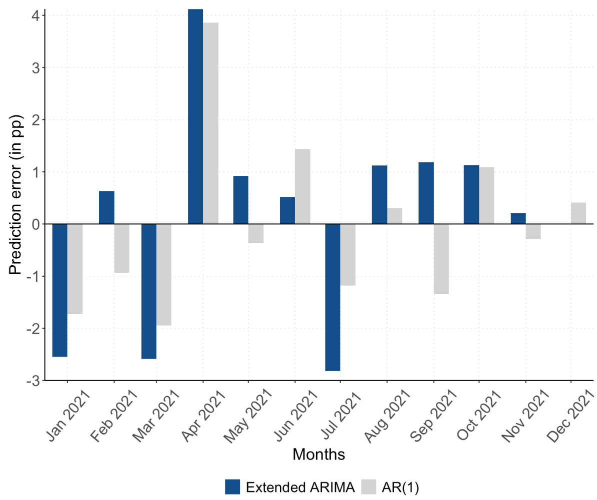
In addition, there does not seem to be an over- or under prediction of inflation for the divisions except COICOP 01. For this division, the predicted values of AR and ARIMA overestimates the inflation for most months.

Figure 6.1: Prediction error (in pp) from CPI ex-post to extended ARIMA and AR(1). Improved divisions out-of-sample.

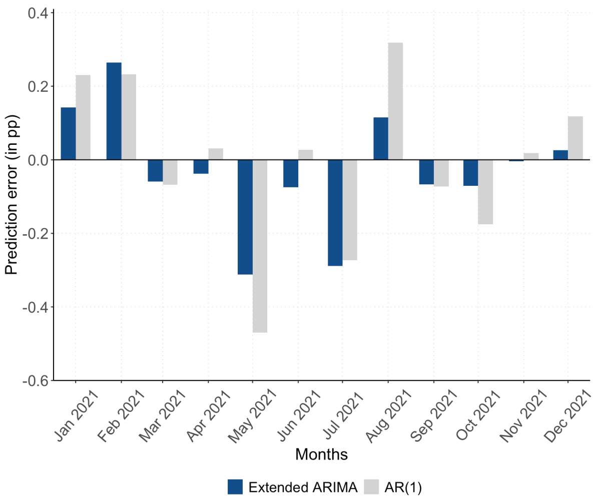
(a) COICOP 01



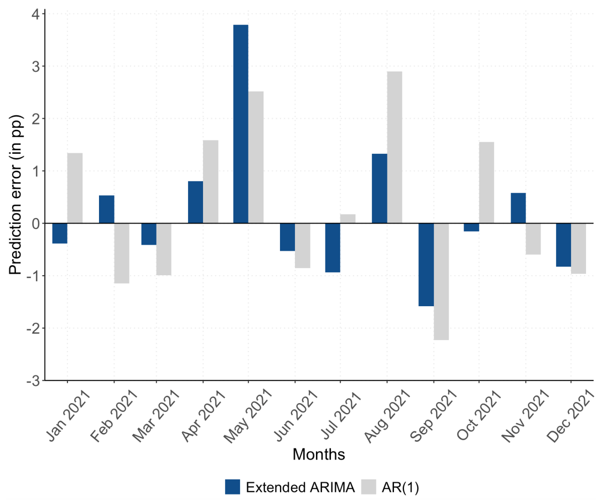
(b) COICOP 03



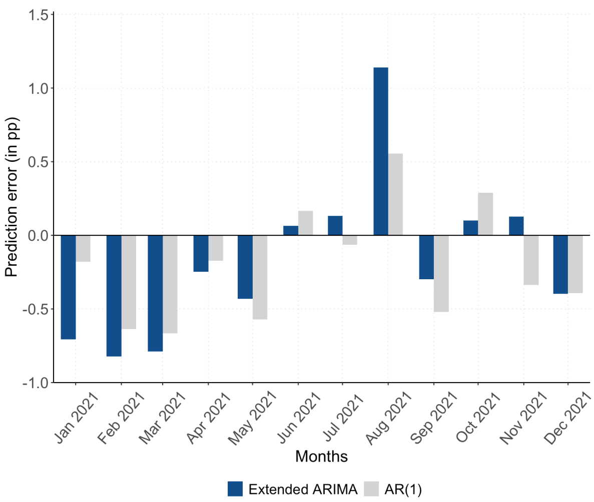
(c) COICOP 04



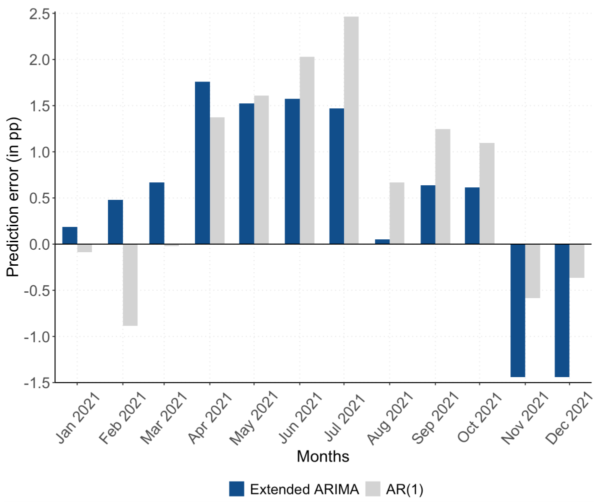
(d) COICOP 05



(e) COICOP 06



(f) COICOP 07



6.3.2 Performance: Trends and Turning Points

Plotting the extended ARIMA and the AR(1) against CPI ex-post facilitates analysing the ability of models to capture short-term developments such as trends and turning points (see Figure 6.2). Moreover, AR(1) models struggle to predict changes. Comparatively, the extended models struggle with similar issues to a smaller extent. Extended models identify trends and turning points more frequently than AR(1).

Extended models outperform their AR(1) benchmarks for all divisions in predicting the correct trend (see Table 6.5). The extended models predict the correct trend in 67% of all months, compared to 52% for the AR(1). Moreover, if we do the same exercise for the models that outperformed the AR(1) in terms of RMSE out-of-sample, compared to the ones that did not, these models predict trends more correctly. The former predicts the correct trend in 73% of all months compared to 57% for the AR(1), while the latter only provides similar results in 55% compared to 41% of all months.

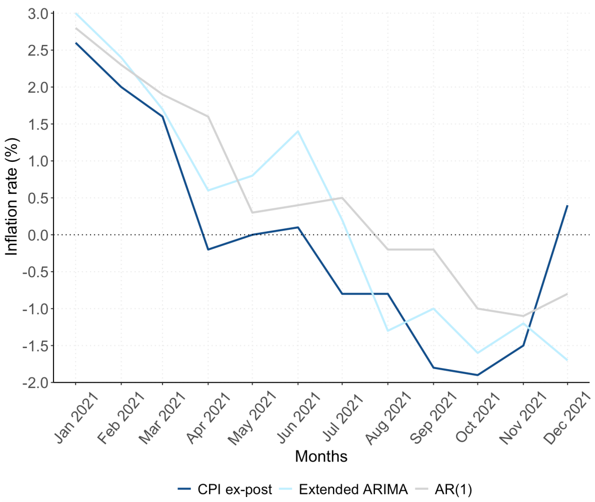
Table 6.5: Correctly predicted trends and turning points extended ARIMA and AR(1) out-of-sample

	Trends		Turning Points	
Division	Extended ARIMA	AR(1)	Extended ARIMA	AR(1)
Improved in terms of RMSE from 6.3.1				
01 Food and non-alcoholic beverages	82%	73%	60%	30%
04 Housing, water, electricity, gas and other fuels	73%	55%	50%	20%
05 Furnishings, household equipment and routine maintenance	55%	36%	20%	0%
07 Transport	82%	64%	60%	40%
Average	**73%**	**57%**	**48%**	**23%**
Unimproved in terms of RMSE from 6.3.1				
03 Clothing and footwear	45%	18%	10%	10%
06 Health	64%	64%	40%	40%
Average	**55%**	**41%**	**25%**	**25%**
Total, all models	**67%**	**52%**	**40%**	**23%**

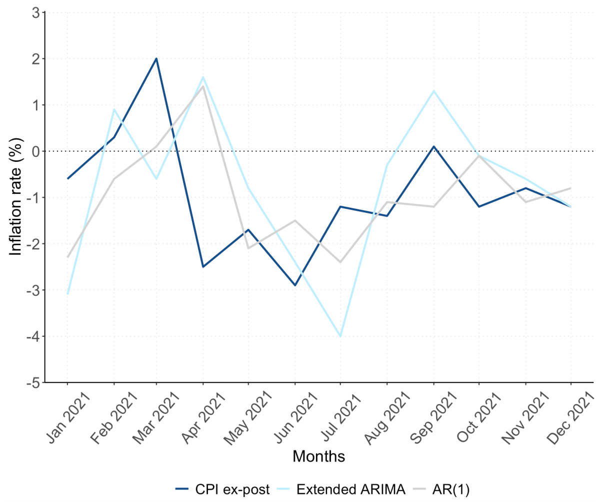
Another feature to consider when assessing the ability to identify the short-term developments is whether the models predict correct turning points. For extended models, 40% of the turning points coincide with the CPI ex-post (see Figure 6.2). In comparison, the AR(1) models predict correctly only 23% of the time. When looking at the four divisions improved by adding Google Indicators, the numbers are 48% compared to 23%. Primarily, the nowcasted series for COICOP 01 and 07 performs well compared to other divisions. Both divisions correctly predict trends and turning points 82% and 60% of the time, respectively. The extended models outperform the AR(1) for all divisions in terms of identifying short-term changes.

Figure 6.2: Nowcasted values for extended ARIMA and AR(1) and CPI ex-post out-of-sample

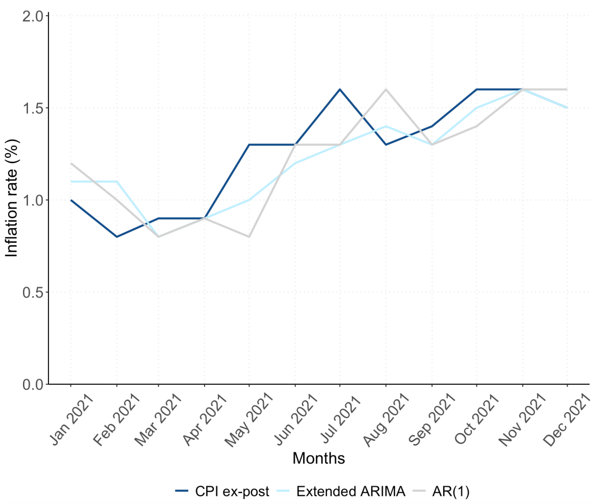
(a) COICOP 01



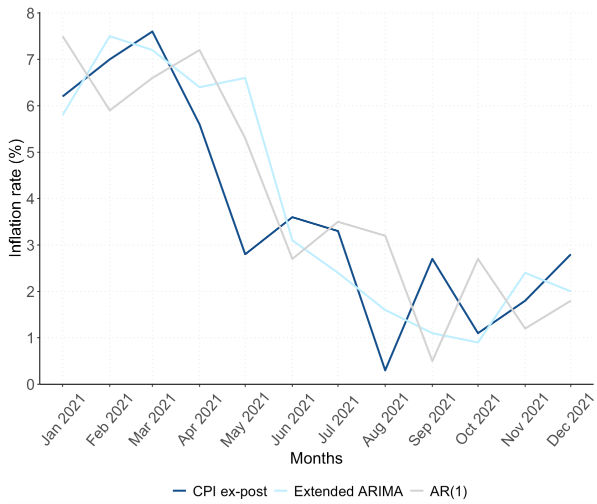
(b) COICOP 03



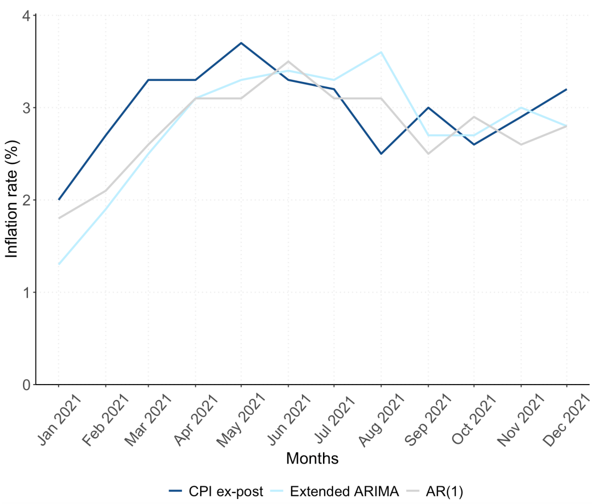
(c) COICOP 04



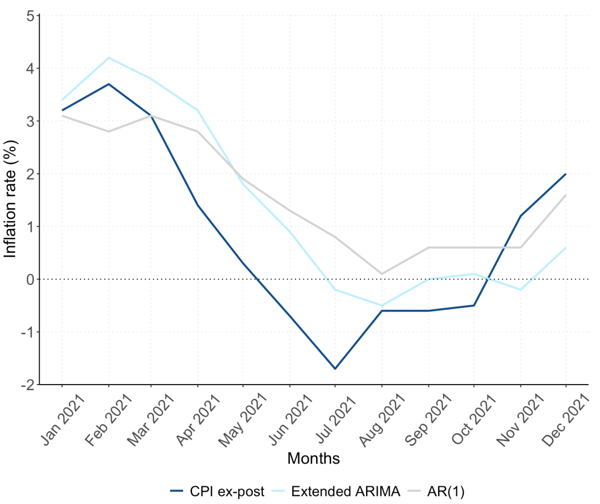
(d) COICOP 05



(e) COICOP 06



(f) COICOP 07



6.4 Summary

Google search queries lower prediction error for all divisions of inflation when included for AR(1) models in Step 1. In addition, for six of twelve divisions, ARIMA models extended with Google Indicators are selected based on in-sample prediction error compared to baseline models in Step 2. When predicting out-of-sample in Step 3, the number of improved divisions is reduced to four, where only COICOP 01, 04, 05 and 07 outperform the AR(1) in terms of RMSE. Further, extended models predict the correct trends and turning points more often than the AR(1) model for six of six divisions.

7 Discussion

In this chapter, we summarise the main findings presented in Chapter 6 to answer our research question presented in Chapter 1. This is answered through three sub-questions presented below. Finally, in Section 7.4, we summarise our findings to answer the main research question:

Can micro behaviour data represented by Google search queries predict divisions of inflation in real-time more precisely than the simple AR(1) process?

7.1 Sub-Question I: In-Sample Model Selection

Can Google search queries improve in-sample model selection for divisions of inflation relative to baseline models from 2011 to 2020?

To answer this sub-question, we analyse the results from the in-sample model selection in Step 2 (see Section 6.2). The major findings are discussed below.

7.1.1 In-Sample Division Analysis

Our major finding is that adding Google search queries to baseline models improve in-sample model selection for six of twelve divisions of inflation. For improved divisions, the positive impact of Google data appears to be greater than the negative impact on the divisions that did not improve. Despite the clear line of separation between improved and unimproved divisions, the distinction between the attributes of their divisions appears to be less clear. In general, divisions consisting of goods are improved more than divisions of services. For the remainder of this sub-question, we discuss the features of the underlying inflation data of the two groups and the impact of their respective Google Indicators.

Improved Divisions

The first common feature we observe for improved divisions is their strong association with goods (see Table 7.1). Of the six improved divisions, only two consist primarily of services, namely housing, water, electricity, gas and other fuels (COICOP 04) and health (COICOP 06). On the other hand, goods are slightly more fragmented as they are further categorised by their durability. In this analysis, we separate between three degrees of durability: durable, semi-durable and non-durable goods.

Table 7.1: Classification of improved CPI divisions. Dark blue indicates majority of consumption type for the division. Light blue the minority.

	Goods			
Division	Durable	Semi-durable	Non-durable	Services
01 Food and non-alcoholic beverages				
03 Clothing and footwear				
04 Housing, water, electricity, gas and other fuels				
05 Furnishings, household equipment and routine maintenance				
06 Health				
07 Transport				

Durable Goods

Google search data improved all divisions of durable goods, namely furnishing, household equipment and routine maintenance (COICOP 05) and transport (COICOP 07). Both divisions have some elements of semi-durable or non-durable goods; however, these are all minor. We propose some plausible reasons why Google search data can improve all divisions of durable goods.

To start with, search behaviour can either represent an intent to purchase goods online or to browse the internet to acquire knowledge about the goods in question (Carrière-Swallow & Labbé, 2013). Durable goods are typically larger investments; therefore, it is more plausible to assume that the Google search data will represent the dimension of acquiring knowledge. Thus, if Google searches are a tool to plan large investments, one can argue that CPI will exhibit a lag compared to the indicator. Moreover, durable goods have historically been more volatile. Both divisions exhibit higher volatility than the average of all divisions. The improved ability to select models for these divisions may be attributed to their volatile nature.

Semi-Durable Goods

The only division comprised of semi-durable goods to be improved by Google search data is clothing and footwear (COICOP 03). Regarding improvement, COICOP 03 experienced the most significant enhancement from Google data (see Table 6.3). This result may have several explanations. For example, category compatibility of the division, the relevance of Google searches, and volatility of the underlying inflation data may explain this result.

First, the division is a composite of strictly semi-durable goods with a specific purpose, making the division more compatible with the relevant Google Category (Apparel). Second, the improvement for this division may be attributed to the quality of the Google Indicator. The uprise and acceleration in e-commerce is a vital contributor to the rapidly increasing number of searches related to apparel (Algolia, 2020). For a division like clothing and footwear, where many purchases are made online, tracking online search behaviour seems to be a positive supplement for the computation of CPI. Thus, if search data on retailers, garments or brands represent a desire to purchase the very moment the search is conducted, then the Google Indicator may be leading to CPI for the clothing and footwear division.

Lastly, the improvement for this division may be attributed to the volatility of the underlying inflation data. Historically, clothing and footwear have been a volatile division. In that sense, any additional information may be of value.

Non-Durable Goods

Similarly to semi-durable goods, Google search data are only able to improve model selection for one division, namely food and non-alcoholic beverages (COICOP 01). However, the improvement of this division may be attributed to less apparent reasons due to its ambiguous construction.

First, the division comprises two distinct components, *"food"* and *"non-alcoholic beverages"*. These components represent separate categories in Google Trends. The compatible Google Category to the first component is selected in the model selection, namely *"Cooking and recipes"*. Nonetheless, *"food"* accounts for 87% of this division. As such, the indicator may capture the most considerable variation of COICOP 01. Second, the selected Google Category does not necessarily represent the demand for foods or groceries, as most purchases are made in physical stores. At best, two arguments can be made to support the selected category. First, demand for food-related activities may unveil demand for groceries. Moreover, Google searches of such characters may be directly linked to demand for groceries through online food delivery suppliers such as Oda (Kolonial) or Adams Matkasse. Nonetheless, isolating the cause of the improvement is difficult due to the construction of the division.

Another explanation for the improvement is that the volatility of the underlying inflation data is above the average of the twelve divisions. This finding supports our hypothesis that Google data improve model selection for more volatile divisions.

Services

In general, Google search data fail to improve model selection for divisions associated with services. However, as stated in the introduction to this section, there are two exceptions: housing, water, electricity, gas and other fuels (COICOP 04) and health (COICOP 06).

First, we note that the construction of these divisions deviates from the majority of other improved models. For instance, COICOP 04 captures rental prices for housing which is arguably a service. However, a considerable fraction of this division measures the price of non-durable goods such as the supply of electricity, gas and fuels. Meanwhile, our selected Google indicator for this division is categorised as *"Real Estate"*, which is arguably a durable good. This difference in durability illustrates the ambiguous nature of service divisions and the carefulness that must be issued when drawing conclusions about which divisions have been improved by Google Indicators. This ambiguity in construction also applies to a majority of unimproved service divisions.

Moreover, the service divisions that did not improve are more volatile. This feature of service divisions weakens one of the hypotheses put forward in this section. By cross-examining the results from the model selection with the volatility of each division, it is clear that this hypothesis does not hold. This result leads us to two possible explanations. On the one hand, our hypothesis of Google data adding more value in model selection for volatile series is invalid. On the other hand, Google search data is useful in model selection. However only suitable for volatile divisions associated with goods.

7.1.2 Summary

In total, adding Google search queries to baseline models improves in-sample model selection for six of twelve divisions of inflation. Further, we find that models that outperformed their baseline by adding Google search queries are either strictly or partly classified as goods. However, due to how these divisions are constructed, some divisions will contain both categories of consumption groups. As we analyse the data at a division level, the intra-division composition prohibits us from concluding if the improved effect of adding Google Indicators is attributed to the goods or services component. However, as none of the improved divisions is solely composed of services, we can conclude that

Google Indicators have a relatively lower impact in selecting models for services. In terms of improving model selection based on volatility, the results are inconclusive.

7.2 Sub-Question II: Out-of-Sample Prediction Error Analysis

Can Google search queries reduce out-of-sample prediction error to CPI ex-post through 2021 compared to the AR(1) model?

To answer this sub-question, we analyse the results from the out-of-sample predictions in Step 3. To validate our results, we compare the performance of the extended ARIMA to the replicated model of the same order and its baseline ARIMA. This comparison is conducted to verify if the performance of the extended ARIMA is attributed to the Google search data or the properties of ARIMA. The major findings are discussed below.

7.2.1 Extended ARIMA vs AR(1)

Results from the nowcasts show that extended ARIMA models outperform AR(1) models in terms of prediction error for the out-of-sample period (see Table 6.4). This result applies to four of six divisions. Adding Google Indicators reduces RMSE for COICOP 01, 04, 05, and 07, which are divisions consisting mainly of goods. The divisions are also more volatile out-of-sample. For COICOP 03 and 06, the AR(1) model achieves lower RMSE than the extended. This finding contradicts Step 2, where adding Google Indicators suggested the opposite.

Consistent with the findings in Section 7.1, predicting changes in inflation for goods has more value in adding Google search queries than for services. The four extended ARIMA models that outperformed the AR(1) baselines mainly or partly consisted of goods.

The second finding is that for COICOP 03 – Clothing and Footwear, the AR(1) model outperforms the extended model in terms of RMSE. We would expect the opposite result as adding Google Indicators to this division in Step 2 lowered estimators of prediction error the most for all divisions. One possible explanation for this surprising finding is that the volatility of CPI ex-post for the in-sample period is significantly higher than for the out-of-sample period (see Figure 3.1). The volatility is 2.9 percentage points for the in-sample period, compared to 0.9 percentage points for the out-of-sample period. This difference is likely due to the Covid-19 pandemic.

As lockdown measures were implemented, stores had to shut down, and demand dropped. In response to this, clothing stores used heavy discounting as a tactic to stimulate demand, leading to a rapid decline in prices (Econsultancy, 2021). These discounts were maintained until lockdown measures were removed. The out-of-sample period does not include the initial decline in inflation originating from these discounts. However, it contains the period where inflation was relatively stable. Right after our out-of-sample period ends, we see a sharp increase in inflation with greater price volatility as restrictions were lifted. It appears that the volatility out-of-sample is not in line with the full-sample volatility.

This unrepresentative out-of-sample period has two main implications. First, as the AR(1) model uses the lagged values to nowcast inflation, the out-of-sample series for COICOP 03 will likely yield an accurate prediction from such model. As the volatility drastically declines, our estimated models from Step 2 are not the best fit for nowcasting this division out-of-sample. Second, this instability between sample periods indicates that Google search data reduce prediction error to a greater extent for more volatile periods. Thus, Google data is less valuable out-of-sample when volatility decreases.

Abrupt changes originating from new information, which affects inflation, are accounted for by adding Google search queries. In addition, the search queries are likely to reflect the current changes in inflation as it indicates changes in demand for specific products in real-time. In contrast, AR(1) only uses lagged values and do not incorporate real-time micro behaviour that may lead to changes in inflation.

7.2.2 Extended vs Replicated ARIMA Models

To ensure if the improved performance of the extended models is a result of adding Google search queries, not the features of the ARIMA models, we replicate an ARIMA of the same order. This model is of identical order as the extended ARIMA for each month; however, it does not include Google Indicators. Subsequently, we compare the nowcast performances. If there are no significant differences in RMSE, then the improved performance compared to the AR(1) model is likely attributed to the features of the ARIMA model, not the Google indicators.

The replicated models outperform the extended models in four of six divisions in terms of RMSE (see Appendix A6; Figure 7.1). Extended models only achieve marginally lower RMSE for COICOP 03 – Clothing and COICOP 04 - Household. In contrast to the findings from previous sections, the results show that Google search queries increase the prediction error when included. This result goes against the hypothesis that Google search queries can improve nowcasting performance in terms of reduced prediction error. These results are in contrast to similar research (Li, Shang, Shouyang, & Ma, 2015; Samantha, 2019).

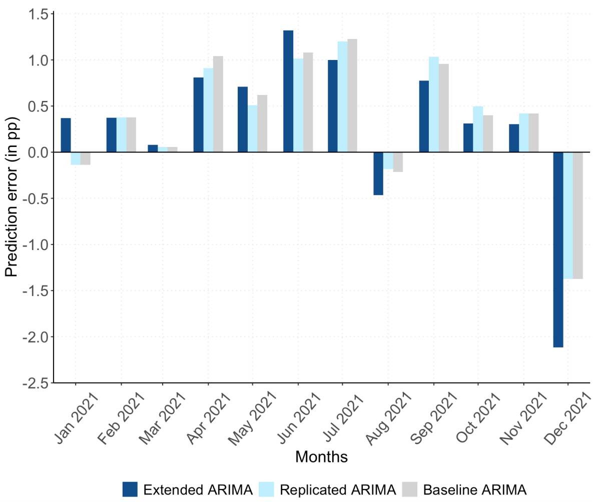
7.2.3 Baseline vs Replicated ARIMA Models

As a final analysis, we assess if the Google Indicators contribute to selecting the optimal ARIMA models. This analysis is performed by reintroducing the baseline ARIMA model from Step 2. We compare the RMSE for the replicated ARIMA against the ARIMA baseline. If the replicas outperform their respective baselines, this indicates that Google search queries provide value in improving the model selection. In particular, it contributes to selecting orders of ARIMA, which improves predictive ability.

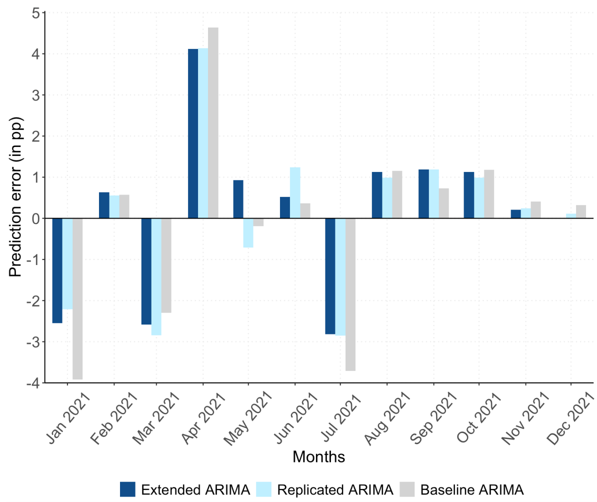
Replicated models outperform the baselines for five of six improved divisions in terms of RMSE (see Appendix A6; Figure 7.1). The only exception is COICOP 06, where the baseline model has a marginally lower RMSE. Consequently, there seems to be a value in including Google search queries in the model selection for the ARIMA, as our findings indicate lower forecasting error.

Figure 7.1: Prediction error (in pp) from CPI ex-post to extended ARIMA, replicated ARIMA, and baseline. Improved divisions out-of-sample.

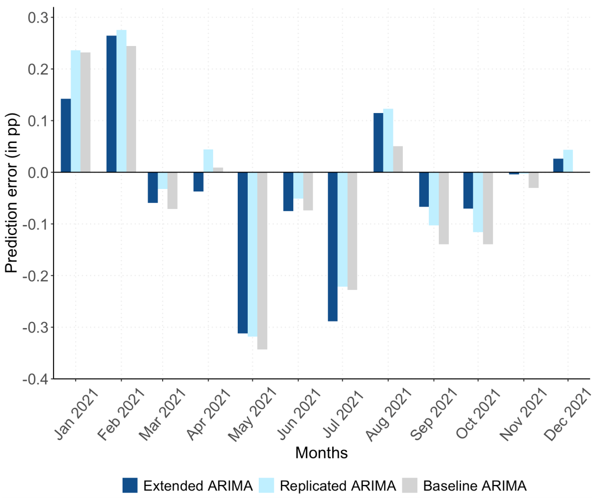
(a) COICOP 01



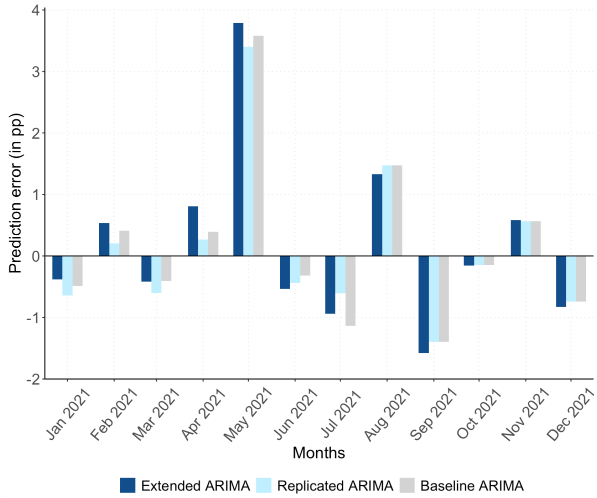
(b) COICOP 03



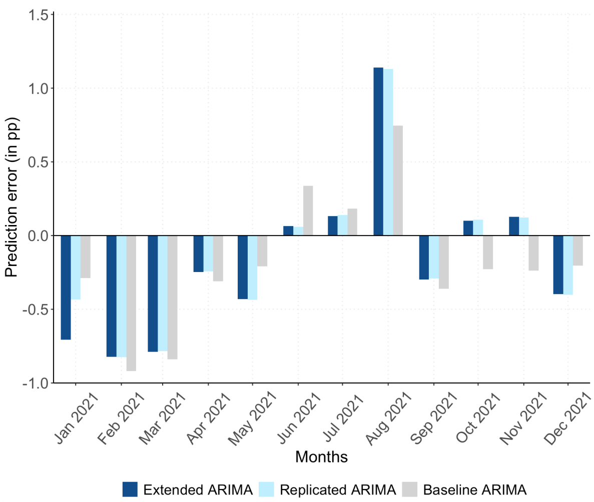
(c) COICOP 04



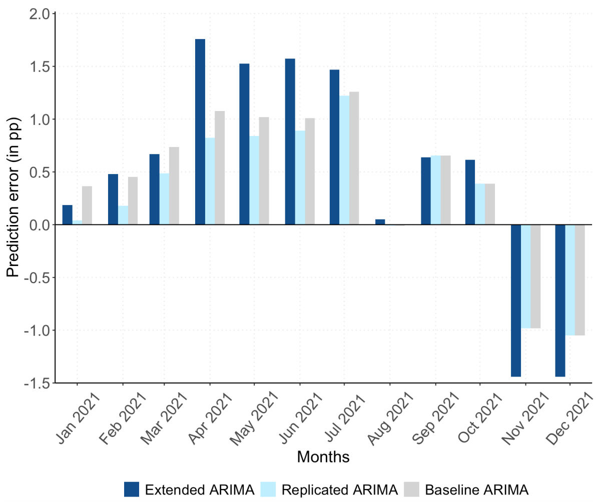
(d) COICOP 05



(e) COICOP 06



(f) COICOP 07



7.2.4 Summary

Extended models reduce the prediction error compared to the AR(1) through 2021 for out-of-sample nowcasts. This finding applies to four of six divisions. These are divisions consisting mainly or partly of goods, which illustrates a volatile time series of inflation. However, the improvement originates from the properties of ARIMA models, not the Google search queries. The only value of adding Google search queries is in reducing RMSE is in the model selection.

7.3 Sub-Question III: Out-of-Sample Trend and Turning Point Analysis

Can Google search queries improve the out-of-sample ability to predict short-term changes in inflation through 2021 compared to the AR(1) model?

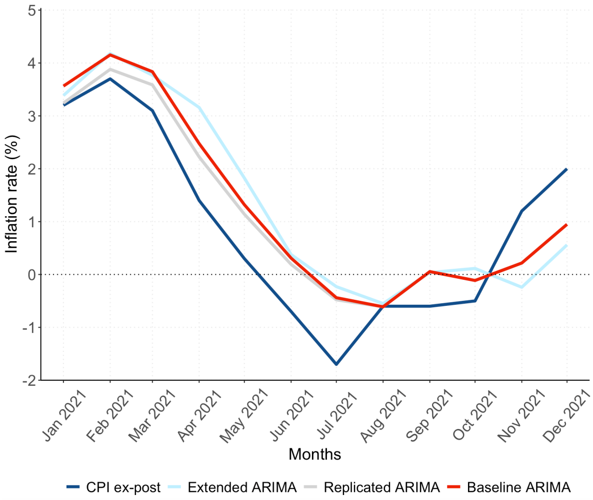
7.3.1 Extended ARIMA vs AR(1)

Our first finding is that extended models consistently predict trends and turning points more correctly than AR(1) for all divisions. Consequently, the extended model outperforms the AR(1) in identifying short-term changes in inflation. Adding Google search data appears to capture micro behaviour in real-time, which the AR(1) does not.

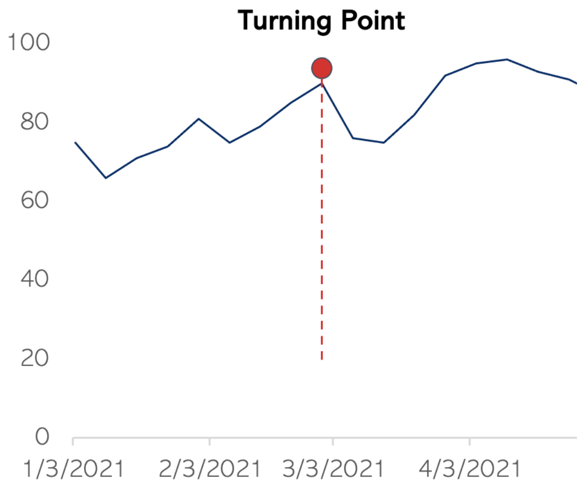
One such example is illustrated in Figure 7.2. The extended ARIMA identifies the turning point in COICOP 07 – Transport, in February 2021, where the rise in inflation turned and started to increase at a lower rate. However, AR(1) fails to capture this turning point before the following month. Figure 7.2 depicts the relative number of Google searches for *"Price"* in the category *"Autos & Vehicles"* for February 2021. This peak might represent a change in micro behaviour as fewer conducted price related searches for vehicles after this month. As a result, it may represent lower demand and a drop in the inflation of vehicles (COICOP 07).

Figure 7.2: Nowcasted values COICOP 07 extended ARIMA and AR(1) and CPI ex-post (left). Google Trends search query for keyword "price" filtered by Google Category "Autos Vehicles" in 2021 (right)

(a) COICOP 07



(b) Google Trends



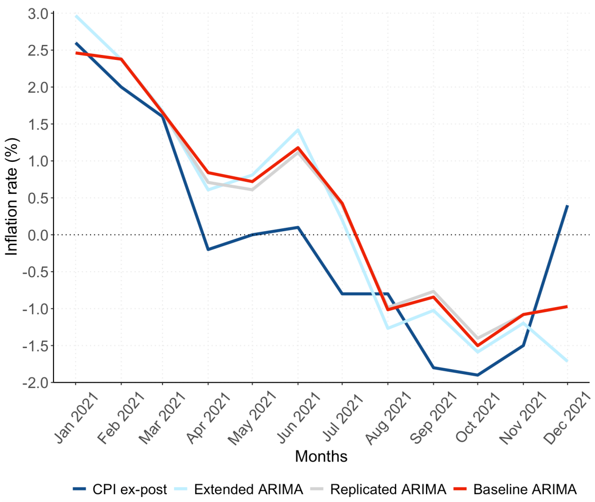
This feature of the extended models can help solve the issue of official inflation data being released with a lag of ten days. One of the drawbacks of this lag is that abrupt changes are not identified, and decision-makers may implement policies on incomplete data. By rapidly identifying the change in trend, more precise and effective decisions can be implemented. Including Google search data can close the gap of the ten days between the release of Google data and the official data.

7.3.2 Extended vs Replicated ARIMA Models

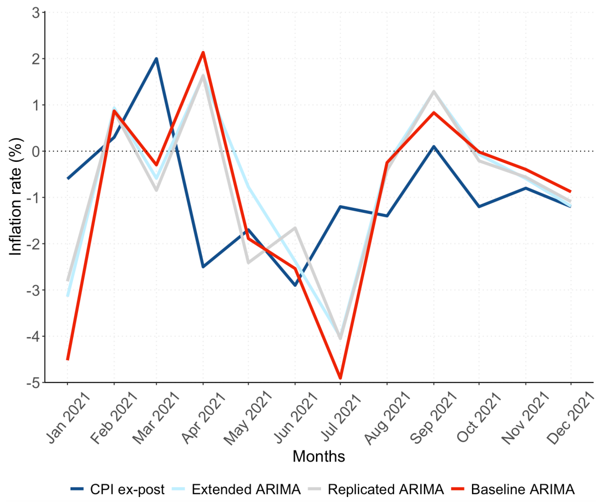
To validate if the performance of extended models originates from Google search data or the properties of ARIMA, they are compared to corresponding replicas. Extended models perform marginally better in predicting correct trends and turning points (see Figure 7.3). For five of six models, they predict the same number of correct trends and turnings points. The differences are COICOP 01 for trends and COICOP 03 for turning points (see Appendix A6). For COICOP 01, the extended model predicts 60% of the turning points compared to 50% for the replica. While for COICOP 03, the extended model predicts the correct trend in 45% of all cases, against 36% for the replica. These differences are minor, and there seems to be little evidence of Google search queries improving nowcasts of inflation in terms of identifying short-term changes.

Figure 7.3: Nowcasted values for extended ARIMA, replicated ARIMA, baseline, and CPI ex-post out-of-sample

(a) COICOP 01

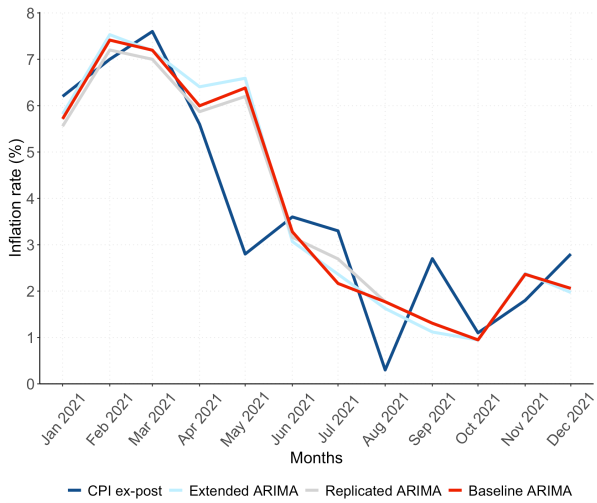


(b) COICOP 03

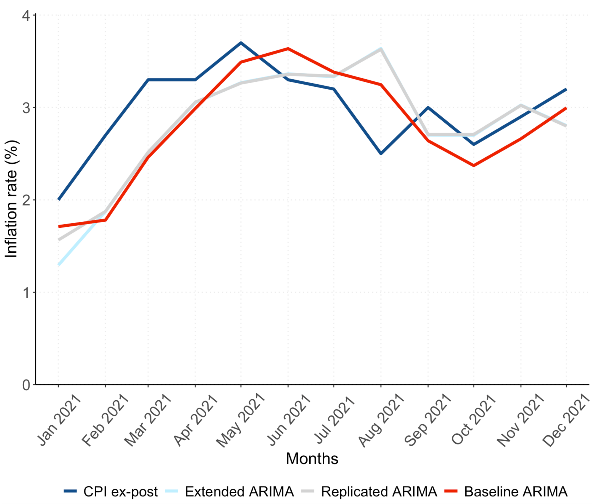


(c) COICOP 04

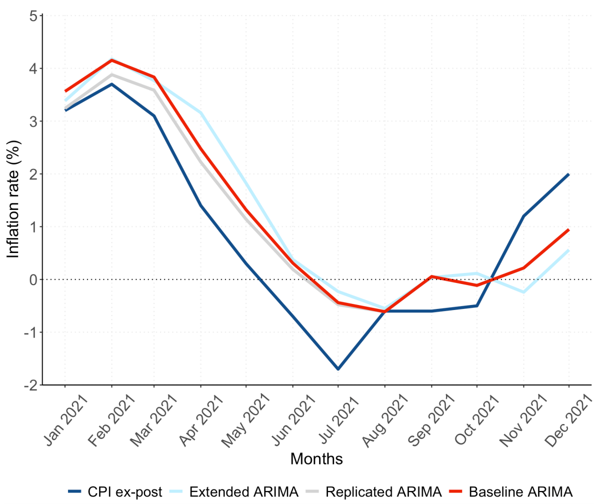
(d) COICOP 05



(e) COICOP 06



(f) COICOP 07



7.3.3 Baseline vs Replicated ARIMA Models

When analysing if Google search queries improve the model selection of optimal ARIMA orders, we see that the baseline and the replica marginally differ in performance (see Appendix A6). The baseline model performs similarly for three of six divisions in terms of predicting correct trends and turning points. Moreover, the baseline outperforms its counterpart model for the remaining three divisions. These results invalidate the conclusion from Section 7.3.1 regarding Google search data improving models' ability to identify short-term changes for the out-of-sample period of 2021. Including Google Indicators reduces the performance of ARIMA models. The properties of ARIMA models are the determining factor in detecting short-term changes better than AR(1).

7.3.4 Summary

Extended models consistently outperform AR(1) in identifying trends and turning points through 2021 for out-of-sample nowcasts. This result applies to all six divisions. However, the improvement originates from the properties of ARIMA models rather than the Google search queries. There does not seem to be a value in incorporating real-time micro behaviour data, represented by Google Trends, in predicting short-term inflation changes.

7.4 Research Question

Can micro behaviour data represented by Google search queries predict divisions of inflation in real-time more precisely than the simple AR(1) process?

From the sub-questions above, we observe that ARIMA models extended with Google search queries outperform the simple AR(1) process for real-time predictions of inflation in Norway through 2021. From Section 7.1, we observe an improved predictive ability in prediction error for models incorporating Google data in six of twelve divisions in-sample. Further, when we use this result to predict divisions of inflation out-of-sample in Section 7.2, prediction errors are reduced for four of six divisions. Moreover, Section 7.3 concludes that extended models improve all divisions in predicting short-term changes. However, the value of adding Google search data appears to be in the model selection. The improved performance is attributed to the properties of ARIMA rather than the predictive ability of Google Indicators.

7.5 Limitations

7.5.1 Quality of Data

Google data has multiple technical and theoretical challenges, which limits this thesis. The first limitation is the poor quality of data between 2004 and 2011. The data from this period is unstable for two reasons. First, the search volume is low and volatile. Second, the structural break in 2011 may cause issues combining the data before and after. Without these issues, we could have included more observations and investigated changes in inflation from 2004 to date. For instance, analysing changes during and after the financial crisis in 2008 would have been of interest as there were abrupt changes in inflation (Gjedrem, 2009).

Several CPI divisions are not split into exclusive categories of goods and services. This makes it challenging to identify how much goods or services benefit from adding Google data. Further analysis of lower-level hierarchical divisions of inflation and Google categories should be conducted to solve this issue. However, it is challenging to identify Google categories that correspond to lower-level inflation divisions and have a sufficiently stable and representative time series. Lower hierarchical levels of Google categories are susceptible to low and volatile search volumes. This issue originates from the nature of how Google categories are constructed. The further down the hierarchy, the fewer search queries and keywords are captured in each category. Thus, we cannot use lower-level divisions and categories of CPI and Google categories to conduct this analysis.

Another limitation is that our in-sample and out-of-sample periods differ in terms of statistical properties. For instance, the variance of the two periods is significantly different. This difference is likely due to our out-of-sample period containing the Covid-19 pandemic. The service sector experienced a considerable reduction in activity, which may have affected the price of goods. As a result, the estimated models from the in-sample period are unlikely to yield the optimal model for predicting out-of-sample. This can explain why the results from the in-sample model selection in Step 1 and Step 2 differ from the results from the out-of-sample predictions in Step 3.

7.5.2 Micro Behaviour Data

We have assumed that Google search queries reflect the public opinion on inflation. However, there might be other motivations behind each search query. For instance, our intuition behind including *"price"* as a keyword is that when individuals search for price changes, they are interested in buying a product. This propensity should then increase demand and then again inflation. Instead, another motivation might be a curiosity about price changes and not increased demand for the product. It is hard to confirm the intention behind every search as we would need insight into the thought process of each individual when the search was conducted. Therefore, it is uncertain if there is a direct link between each specific keyword and inflation. There might be spurious relationships as our included variables might not truly measure inflation.

When using web queries, we do not capture the behaviour of all individuals and firms. Individuals also gather information by going directly to websites instead of using search engines. For instance, when interested in the price of Nike Air Jordan shoes, individuals might go directly to the website of Nike. Google Trends does not capture such search behaviour, which could create bias in our data.

7.5.3 Features of Google Trends

Another limitation is the choice of keywords. Our method for selecting keywords is based on five subjective criteria. This approach has no solid theoretical or empirical foundation as the research in this area is limited. We have assessed the reliability and validity of each Google search query to make up for this limitation. However, we recognise that keywords better fit to explain changes in inflation may have been excluded due to our selection process.

The construction process of Google topics and categories is not publicly available for scrutiny, making it difficult to verify that the divisions of inflation and Google categories overlap. Thus, we cannot be sure that the Google categories added to the models coincide with respective divisions. Further, we cannot be sure that the data explain the variation in inflation. Even though we analysed the sub-levels of Google categories and matched them to sub-levels of divisions of CPI, the allocation process of keywords and search queries is not available. Consequently, we cannot be sure which keywords are in each Google category. This lack of validation may lead to weakly correlated divisions and Google Indicators.

7.5.4 Applied Models

We have limited our research design to models which require included time series to be of the same frequencies. Unfortunately, this limitation prohibits us from using daily or weekly Google search data, as official inflation data from Norway is unavailable at corresponding frequencies. However, some models allow for using mixed frequencies, which could be applied (See Section 2.2).

Even though ARIMA models have proven to forecast adequately, it has received critique for lacking autonomy related to changes in policy. In general, the ARIMA models perform inadequately when forecasting a turning point unless it happens near the long-run equilibrium of the time series (Lucas, 1976).

In step 3, we excluded six divisions of inflation from being nowcasted. We cannot be sure that even though they did not improve by adding Google Indicators in step 2, they would perform subpar to the AR(1) in step 3. As a result, we might have excluded interesting findings as a result of our scope.

7.6 Future Research

This thesis evaluates the predictive power of Google data by assessing prediction and the ability to predict trends and turnings points. Real-time data were applied to close the ten-day gap to which inflation data are subject. However, nowcasts of inflation might be of utility for more extended time horizons and other forecasting purposes. Thus, metrics such as variance and skewness, among others, of the nowcasts might be of interest. These other metrics of forecast errors are beyond the scope of this thesis; however, it creates a foundation for further research.

As mentioned in the limitations, AR and ARIMA models prohibit us from using higher frequencies than monthly observations. However, future research into inflation nowcasts using weekly Google data through mixed frequency models might be of academic interest.

For instance, Li, Shang, Shouyang, & Ma (2015) find in their paper that weekly Google data improve the forecast performance of inflation.

Another area which can be of interest for future research is to look at different forecast horizons. We have limited our thesis to look at short-term changes performing one-month rolling predictions through 2021. Evaluating if the performance of forecasting models predicting inflation using Google search queries improves for either quarterly, 6-month, or yearly horizons may be an exciting area.

To some extent, our thesis investigates if Google search queries might be of more value in predicting goods rather than services. We briefly make a few observations based on our results. However, these findings do not imply causality. The study of what classification of consumption and divisions of inflation benefits most by adding Google search queries is outside this study's scope. Further research requires more specific research on different classifications of consumption—for instance, goods of different durability.

Google Trends does have the feature of limiting the search queries to specific regions in each country. This feature facilitates studying regional changes in inflation. Such models may be useful in detecting regional macroeconomic shocks for different divisions. One such study could be to see if Google search queries managed to identify the price shock in housing prices in the Rogaland region in 2015, which was a result of the oil price shock. Currently, official data is not released for different regions. Google search data could fill this gap in predictions.

Another extension to our thesis could be related to specific events, such as abrupt changes during times of crisis. For instance, inflation is rapidly rising due to over-stimulative fiscal and monetary policies as a response to the Covid-19 pandemic. Thus, analysing if models using Google search data manage to identify spikes in inflation might be of great utility.

Beyond Google Trends, there are other real-time data sources which could be applied in nowcasting. Such an example is Google Mobility data which shows the movement of people for specific locations. This data could be combined with Google Trends to increase the precision of forecast for specific regions or locations.

8 Conclusion

In this thesis, we have examined the feasibility of applying micro behaviour data, represented by Google search queries, to nowcast inflation in Norway. We answered this question by analysing the predictive potential of Google search data across three dimensions. These dimensions are model selection, prediction error and ability to identify short-term changes.

First, we examined the value-added information of Google search data in a preliminary test of content. Then, we analysed which divisions of inflation historically reduced the prediction error the most when adding Google search queries through the Box-Jenkins methodology. Lastly, one-month rolling nowcasts for 2021 were performed for the divisions that were improved in the preceding step. The predictive performances of the extended models were then compared to the simple AR(1) process to analyse prediction errors and the ability to identify short-term changes.

This thesis finds that ARIMA models extended with Google search queries outperformed their benchmarks across all three dimensions. However, these improvements were not uniformly found for all divisions. In general, divisions consisting of goods and demonstrating high volatility benefit most from adding Google search queries. These findings are consistent with model selection, prediction error, and short-term changes, indicating that the inclusion of micro behaviour data captures real-time changes in inflation in contrast to the AR(1) model.

Google data captures real-time changes and micro behaviour. Thus, it has the potential to solve the issue of lagged official data on inflation. Incorporating this new data dimension provides policymakers with an additional tool to conduct increasingly more accurate predictions in real-time.

Despite outperforming the AR(1), the extended model is inferior compared to a replicated ARIMA of the same order and its ARIMA baseline. This finding indicates that the improved predictive ability is attributed to the properties of ARIMA rather than Google search data. However, the replicated model outperforms its baseline in terms of prediction error. Consequently, Google search data is of value in model selection. This way, real-time and micro behaviour data is of value. Thus, further research should focus on using higher frequency Google search data to capture real-time changes to a greater extent than in this thesis.

References

Abay, K. A., Tafere, K., and Woldemichael, A. (2020). Winners and losers from covid-19: Global evidence from google search. *World Bank Policy Research Working Paper*, (9268).

Algolia (2000). 40+ stats on e-commerce search and kpis.

Askitas, N. and Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting.

Baker, S. R. and Fradkin, A. (2017). The impact of unemployment insurance on job search: Evidence from google search data. *Review of Economics and Statistics*, 99(5):756–768.

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.

Bjertnæs, G. H. (2021). *COVID-19, tapt verdiskaping og finanspolitikkens rolle: utredning for Koronakommisjonen*. Oslo-Kongsvinger: Statistisk sentralbyrå.

Bjørnland, H. C. and Thorsrud, L. A. (2015). *Applied time series for macroeconomics*. Gyldendal akademisk.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Box, G. E. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332):1509–1526.

Box, G. E. and Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical association*, 70(349):70–79.

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

Chan, N. H. (2011). *Time series: Applications to finance with R and S-Plus*, volume 837. John Wiley & Sons.

Chen, T., So, E. P. K., Wu, L., and Yan, I. K. M. (2015). The 2007–2008 us recession: What did the real-time google trends data tell the united states? *Contemporary Economic Policy*, 33(2):395–403.

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

Choi, Hyunyoung & Varian, H. (2012). Predicting the present with google trends. *Economic record*, 88:2–9.

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.

Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *The journal of finance*, 66(5):1461–1499.

Doerr, S., Gambacorta, L., et al. (2020). Identifying regions at risk with google trends: the impact of covid-19 on us labour markets. Technical report, Bank for International Settlements.

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.

Econsultancy (2021). Using web-based search data to predict macroeconomic statistics.

Ettredge, M., Gerdes, J., and Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, 48(11):87–92.

Favero, C. and Hendry, F. D. (1992). Testing the lucas critique: A review. *Econometric reviews*, 11(3):265–306.

Ferrara, L. and Simoni, A. (2020). When are google data useful to nowcast gdp? an approach via pre-selection and shrinkage. *arXiv preprint arXiv:2007.00273*.

Fondeur, Y. and Karamé, F. (2013). Can google data help predict french youth unemployment? *Economic Modelling*, 30:117–125.

Giannone, D., Reichlin, L., and Small, D. H. (2006). Nowcasting gdp and inflation: the real-time informational content of macroeconomic data releases.

Gjedrem, S. (2009). Experiences with the financial crisis.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., and Watts, D. J. (2010). Predicting consumer behavior with web search. *Proceedings of the National academy of sciences*, 107(41):17486–17490.

Gonzales, F., Jaax, A., and Mourougane, A. (2020). Nowcasting aggregate services trade. a pilot approach to providing insights into monthly balance of payments data.

Greenacre, M. J. (1984). Theory and applications of correspondence analysis.

Guzman, G. (2011). Internet search behavior as an economic forecasting tool: The case of inflation expectations. *Journal of economic and social measurement*, 36(3):119–167.

Hamilton, J. D. (1996). Specification testing in markov-switching time-series models. *Journal of econometrics*, 70(1):127–157.

Hanken, S. and Syse, P. (2020). 20 år med fleksibel inflasjonsstyring: hvordan har fleksibel inflasjonsstyring påvirket inflasjon, konjunkturer og finansiell stabilitet? Master's thesis.

Hyndman, R. (2014). Thoughts on the ljung-box test. *Unpublished paper, Monash University, Victoria, Australia.*

Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice.* OTexts.

Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for r. *Journal of statistical software*, 27:1–22.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.

Kholodilin, K. A., Podstawski, M., Siliverstovs, B., and Bürgi, C. R. S. (2009). Google searches as a means of improving the nowcasts of key macroeconomic variables.

Lai, S.-M., Hsieh, H.-L., and Chang, C.-W. (2005). Preparative separation of γ-oryzanol from rice bran oil by silica gel column chromatography. *Journal of liquid chromatography & related technologies*, 28(1):145–160.

Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The parable of google flu: traps in big data analysis. *science*, 343(6176):1203–1205.

Li, X., Shang, W., Wang, S., and Ma, J. (2015). A midas modelling framework for chinese inflation index forecast incorporating google search data. *Electronic Commerce Research and Applications*, 14(2):112–125.

Linden, A., Adams, J. L., and Roberts, N. (2003). Evaluating disease management program effectiveness: an introduction to time-series analysis. *Disease Management*, 6(4):243–255.

Marcucci, J. et al. (2009). "google it!" forecasting the us unemployment rate with a google job search index. Technical report, University Library of Munich, Germany.

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

Meyler, A., Kenny, G., and Quinn, T. (1998). Forecasting irish inflation using arima models.

Mittnik, S., Fabozzi, F. J., Focardi, S. M., Rachev, S. T., and Jašić, T. (2007). *Financial econometrics: from basics to advanced modeling techniques*. John Wiley & Sons.

Modugno, M. (2013). Now-casting inflation using high frequency data. *International Journal of Forecasting*, 29(4):664–675.

Morgavi, H. (2020). A garch model to now-cast private consumption using google trends data. *OECD*.

Nakamura, E. (2005). Inflation forecasting using a neural network. *Economics Letters*, 86(3):373–378.

Narita, M. F. and Yin, R. (2018). *In search of information: use of google trends' data to narrow information gaps for low-income developing countries*. International Monetary Fund.

Norges Bank (2010). Models for monetary policy analysis and forecasting.

Norges Bank (2020). Inflation.

Norges Bank (2022). Norges bank's monetary policy handbook.

Olsen, Ø. (2015). Oil and the norwegian economy–the challenges ahead.

Olsen, Ø. (2022). Annual address: Economic perspectives.

Pearson, E. S. (1931). The test of significance for the correlation coefficient. *Journal of the American Statistical Association*, 26(174):128–134.

Pindyck, R. and Rubinfeld, D. (1998). Econometric models and economic forecasts/pindyck r.

Pisu, C. and Hwand (2019). Digital platforms and the covid-19 crisis. *OECD*.

Pisu, C. and Hwand (2022). Inflation is caused by supply and demand. *U.S. Chamber of Commerce*.

Quietly (2015). Exploring google trends' explore function: finding keywords and queries. *OECD*.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Samanta, G. (2019). Does google search index help track and predict inflation rate? an exploratory analysis for india. *Reserve Bank of India*.

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

SOEC (2010). European classification of individual consumption according to purpose.

SSB (2017). Ni av ti surfer på nettet hver dag.

SSB (2021). Ict usage in households.

SSB (2022). Consumer price index.

Statista (2021). Market share held by the leading search engines in norway as of june 2021.

Steindel, C., Cecchetti, S. G., and Chu, R. (2000). The unreliability of inflation indicators. *Available at SSRN 716681*.

Stock, J. H. and Watson, M. W. (2008). Phillips curve inflation forecasts.

Stock, J. H. and Watson, M. W. (2017). Twenty years of time series econometrics in ten pictures. *Journal of Economic Perspectives*, 31(2):59–86.

Svensson, L. E. (2000). Open-economy inflation targeting. *Journal of international economics*, 50(1):155–183.

The International Energy Agency (2018). Global ev outlook 2018, towards cross-modal electrification.

United Nations (2018). Classification of individual consumption according to purpose (coicop). *Department of Economic and Social Affairs, Statistical Papers Series M No. 99*.

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

Wheelwright, S., Makridakis, S., and Hyndman, R. J. (1998). *Forecasting: methods and applications*. John Wiley & Sons.

Wilson, G. T. (2016). Time series analysis: Forecasting and control, by george ep box, gwilym m. jenkins, gregory c. reinsel and greta m. ljung, 2015. published by john wiley and sons inc., hoboken, new jersey, pp. 712. isbn: 978-1-118-67502-1.

Woloszko, N. (2020). Tracking activity in real time with google trends.

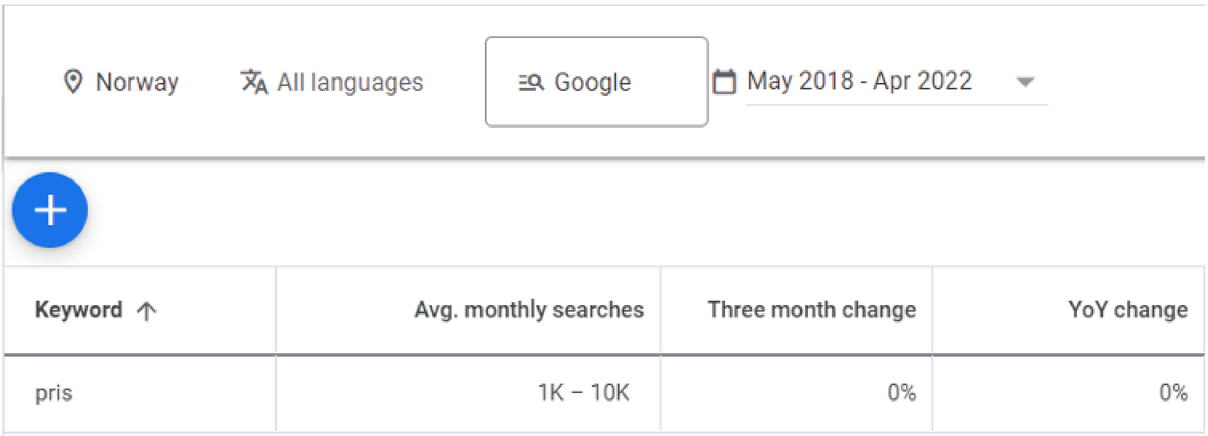
Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

Wu, L. and Brynjolfsson, E. (2015). The future of prediction: How google searches foreshadow housing prices and sales. In *Economic analysis of the digital economy*, pages 89–118. University of Chicago Press.

Appendix

A1 Google Adwords

Figure A1.1: Google Adword – Search Volume Keyword *"Price"*



A2 Divisions of Inflation and Their Respective Weights

Table A2.1: Division Weights: COICOP 01 - 04

			Relative weight	
COICOP	**Division**	**Index**	**of CPI-ATE**	**of division**
00	**CPI-ATE All-Item Index**	**1 000**	**100%**	
01	**Food and non-alcoholic beverages**	**128.6**	**13%**	**100%**
01.1	Food	112.3		87%
01.2	Non-alcoholic beverages	16.3		13%
02	**Alcoholic beverages and tobacco**	**42**	**4%**	**100%**
02.1	Alcoholic beverages	26.4		63%
02.2	Tobacco	15.6		37%
03	**Clothing and footwear**	**50.9**	**5%**	**100%**
03.1	Clothing	43.6		86%
03.2	Footwear	7.2		14%
04	**Housing, water, electricity, gas and other fuels**	**246.2**	**25%**	**100%**
04.1	Actual rentals for housing	44.9		18%
04.2	Imputed rentals for housing	138.5		56%
04.3	Maintenance and repair of the dwelling	1.5		1%
04.5	Water supply and miscellaneous services relating to dwelling	17.5		7%
04.6	Electricity, gas and other fuels	43.8		18%

Table A2.2: Division Weights: COICOP 05 - 08

			Relative weight	
COICOP	**Division**	**Index**	**of CPI-ATE**	**of division**
00	**CPI-ATE All-Item Index**	**1 000**	**100%**	
05	**Furnishings, household equipment and routine maintenance**	**68.9**	**7%**	**100%**
05.1	Furniture and furnishing, carpets and other floor coverings	25.2		38%
05.2	Household textiles	8.5		12%
05.3	Household appliances	10.5		15%
05.4	Glassware, tableware and household utensils	7.7		11%
05.5	Tools and equipment for house and garden	10.4		15%
05.6	Goods and services for routine household maintenance	6.5		9%
06	**Health**	**33.9**	**3%**	**100%**
06.1	Medical products, appliances and equipment	15.9		49%
06.2	Out-patient services	17.2		51%
07	**Transport**	**149.7**	**15%**	**100%**
07.1	Purchase of vehicles	68.2		46%
07.2	Operation of personal transport	56.2		38%
07.3	Transport services	25.3		16%
08.1	Postal services	0.7		3%
08.2	Telephone equipment	5.8		24%
08.3	Telephone services	17.3		73%
08	**Communications**	**23.7**	**2%**	**100%**
08.1	Postal services	0.7		3%
08.2	Telephone equipment	5.8		24%
08.3	Telephone services	17.3		73%

Table A2.3: Division Weights: COICOP 09 - 12

			Relative weight	
COICOP	**Division**	**Index**	**of CPI-ATE**	**of division**
00	**CPI-ATE All-Item Index**	**1 000**	**100%**	
09	**Recreation and culture**	**107.9**	**11%**	**100%**
09.1	Audio-visual, photographic and information processing equipment	22.4		21%
09.2	Other major durables for recreation and culture	7.4		6%
09.3	Other recreational items and equipment, gardens and pets	23.3		22%
09.4	Recreational and cultural services	28.7		27%
09.5	Newspapers, books and stationery	18.3		17%
09.6	Package holidays	7.7		7%
10	**Education**	**4.9**	**0%**	**100%**
11	**Restaurants and hotels**	**56.2**	**6%**	**100%**
11.1	Restaurant services	48.3		86%
11.2	Accommodation services	7.9		14%
12	**Miscellaneous goods and services**	**87.2**	**9%**	**100%**
12.1	Personal care	30.5		35%
12.2				
12.3	Personal effects n.e.c.	4.1		4%
12.4	Social protection	16.6		19%
12.5	Insurance	15.5		18%
12.6	Financial services n.e.c.	12.9		15%
12.7	Other services n.e.c.	7.6		9%

A3 Selected Google Categories by Hierarchical Level

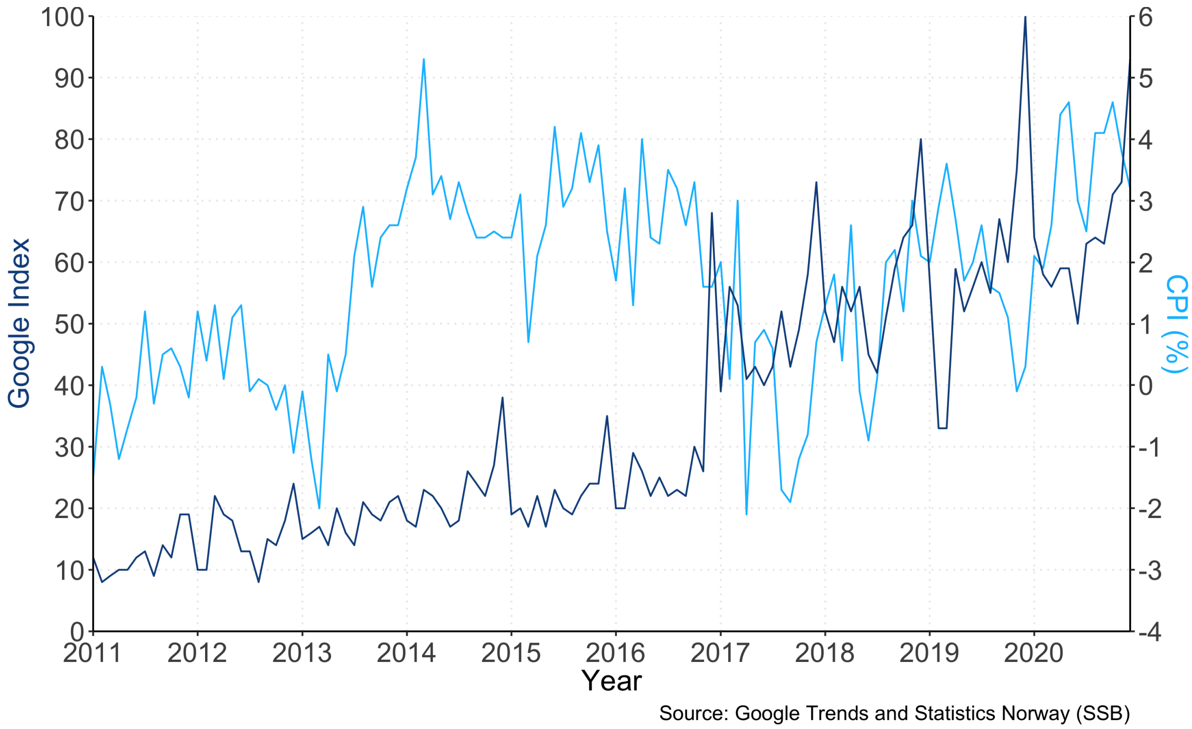
Table A3.1: Selected Google Categories by Hierarchical Level

COICOP	Category 1	Level 1	Category 2	Level 2
01	Cooking & Recipes	2	Non-Alcoholic Beverages	2
02	Alcoholic Beverages	2	Tobacco Products	2
03	Apparel	2		
04	Real Estate	1		
05	Home & Garden	1		
06	Health	1		
07	Autos & Vehicles	1	Transportation & Logistics	2
08	Internet & Telecom	1		
09	Hobbies & Leisure	1		
10	Education	2		
11	Hotels & Accommodations	2	Restaurants	2
12	Beauty & Fitness	1	Social Services	2

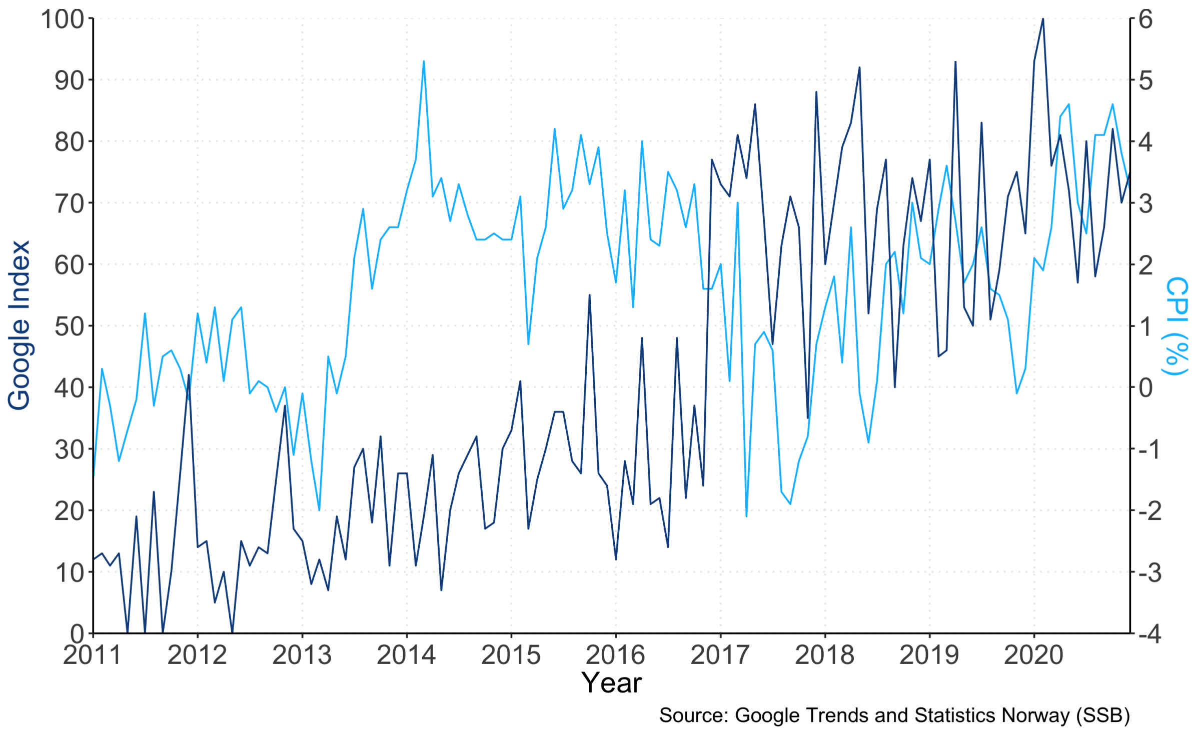
A4 Plots: COICOP vs Google Indicator

A4.1 COICOP 01 – Food and Non-alcoholic Beverages

(a) GI: Cooking & recipes

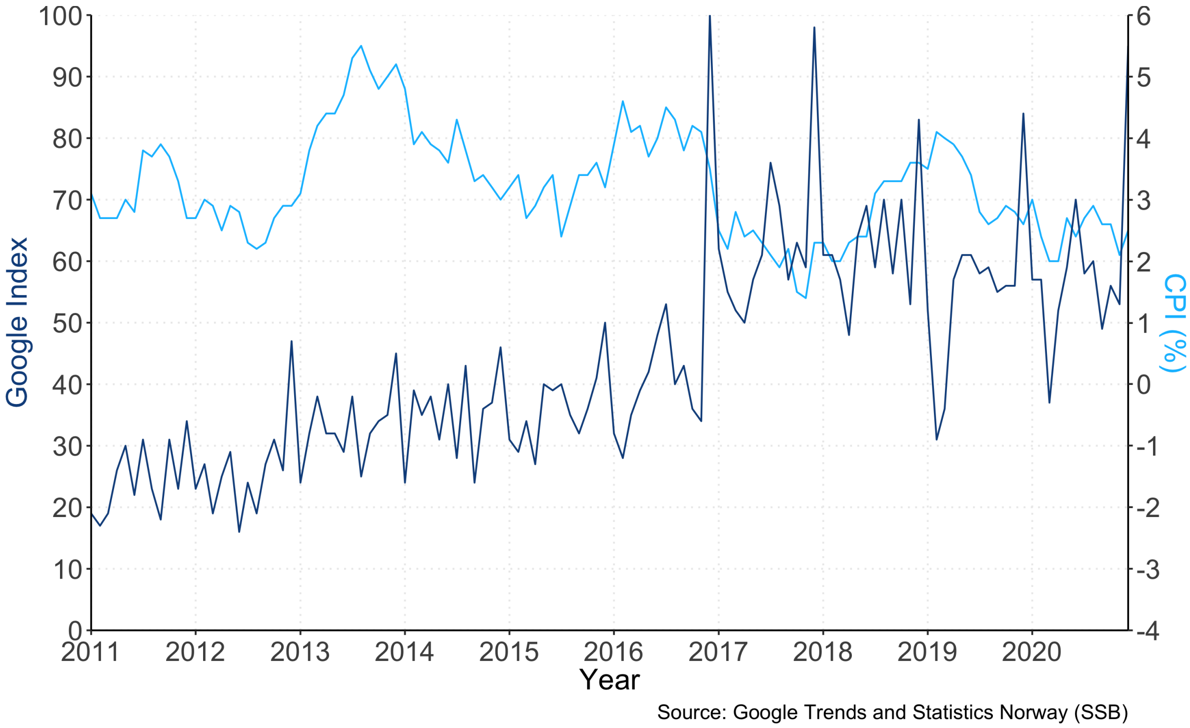


(b) GI: Non-alcoholic beverages

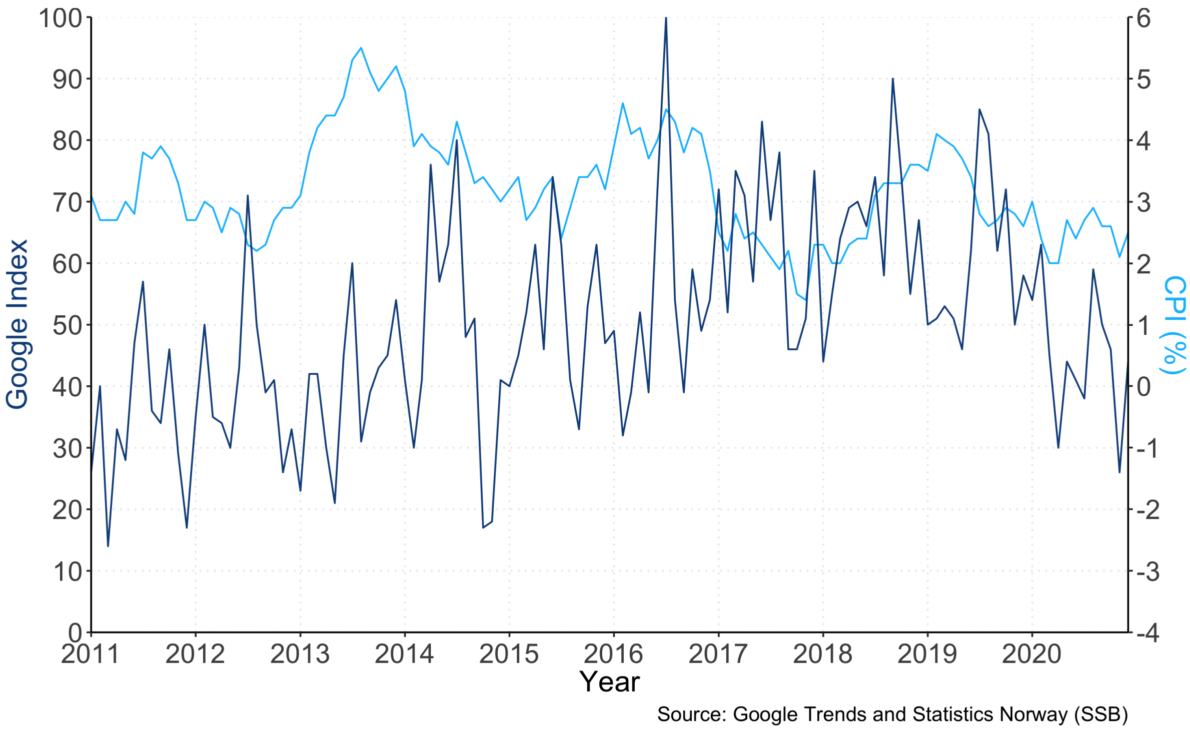


A4.2 COICOP 02 – Alcoholic Beverages and Tobacco

(a) GI: Alcoholic beverages

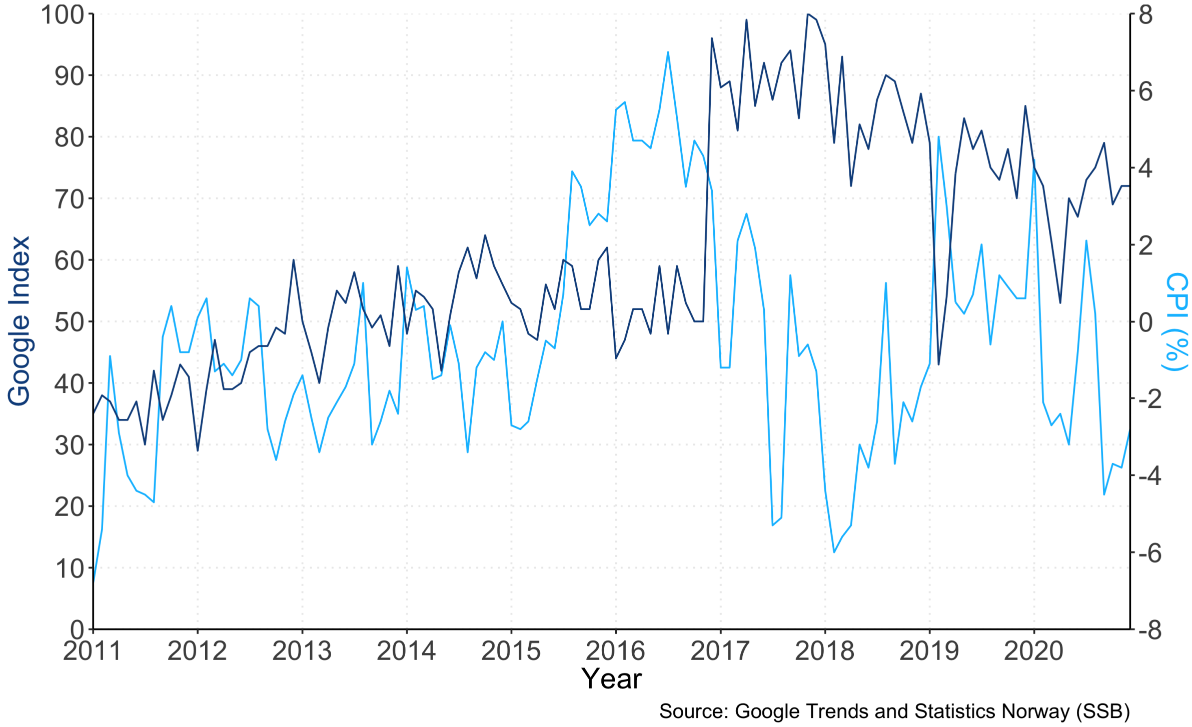


(b) GI: Tobacco



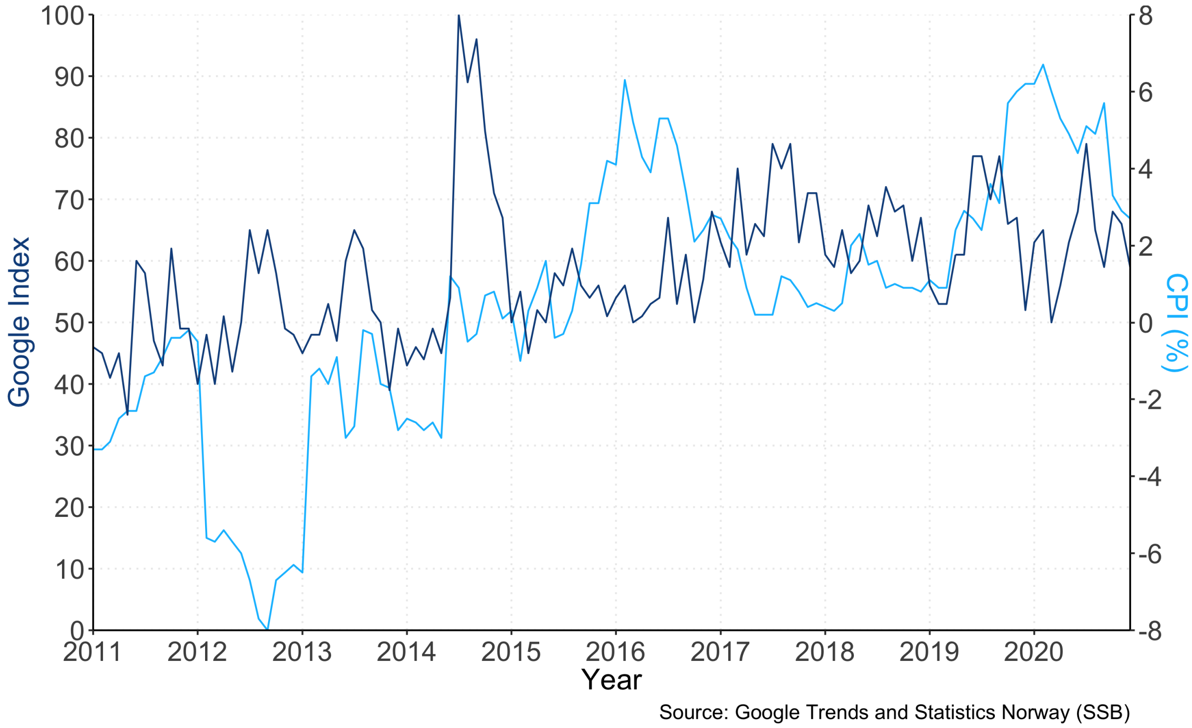
A4.3 COICOP 03 – Clothing and Footwear

Figure A4.3: GI: Apparel



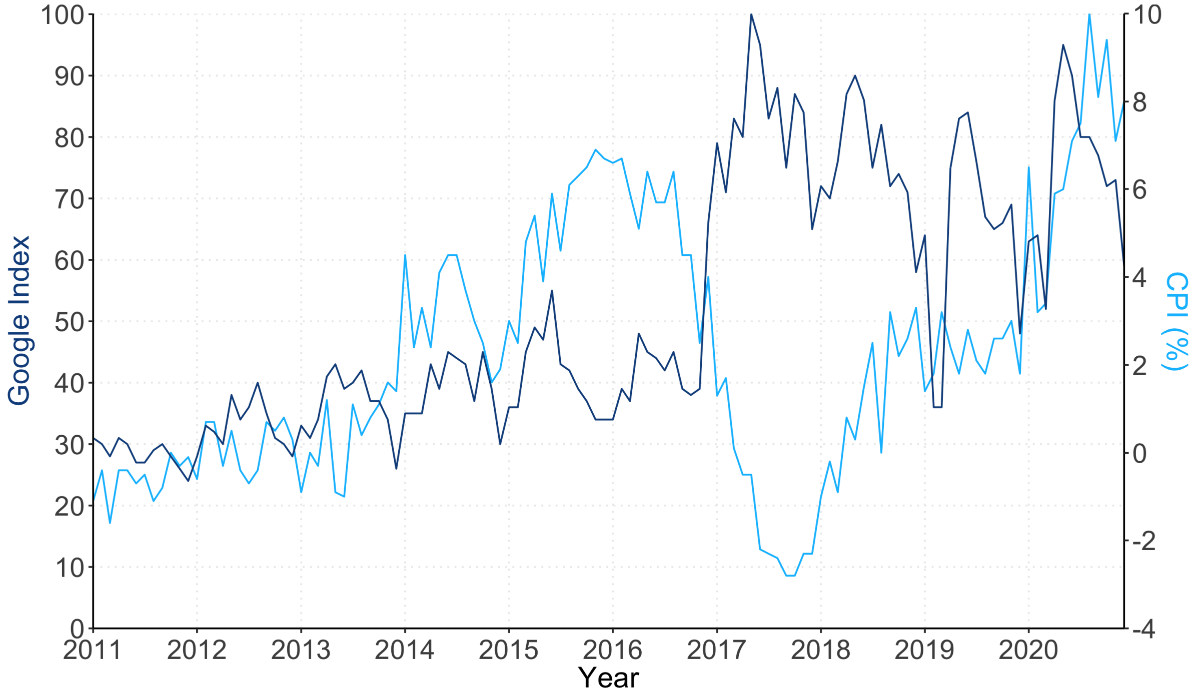
A4.4 COICOP 04 – Housing, Water, Electricity, Gas and Other Fuels

Figure A4.4: GI: Real Estate



A4.5 COICOP 05 – Furnishings, Household, and Routine Maintenance

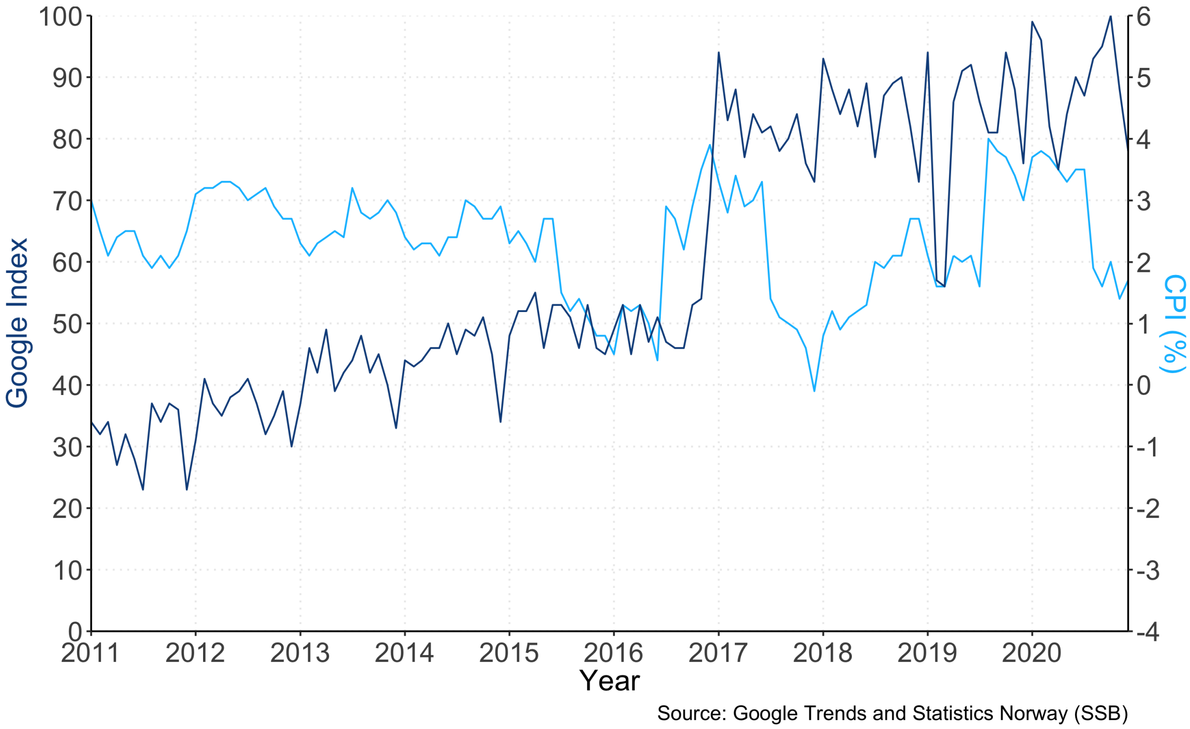
Figure A4.5: GI: Home & garden



Source: Google Trends and Statistics Norway (SSB)

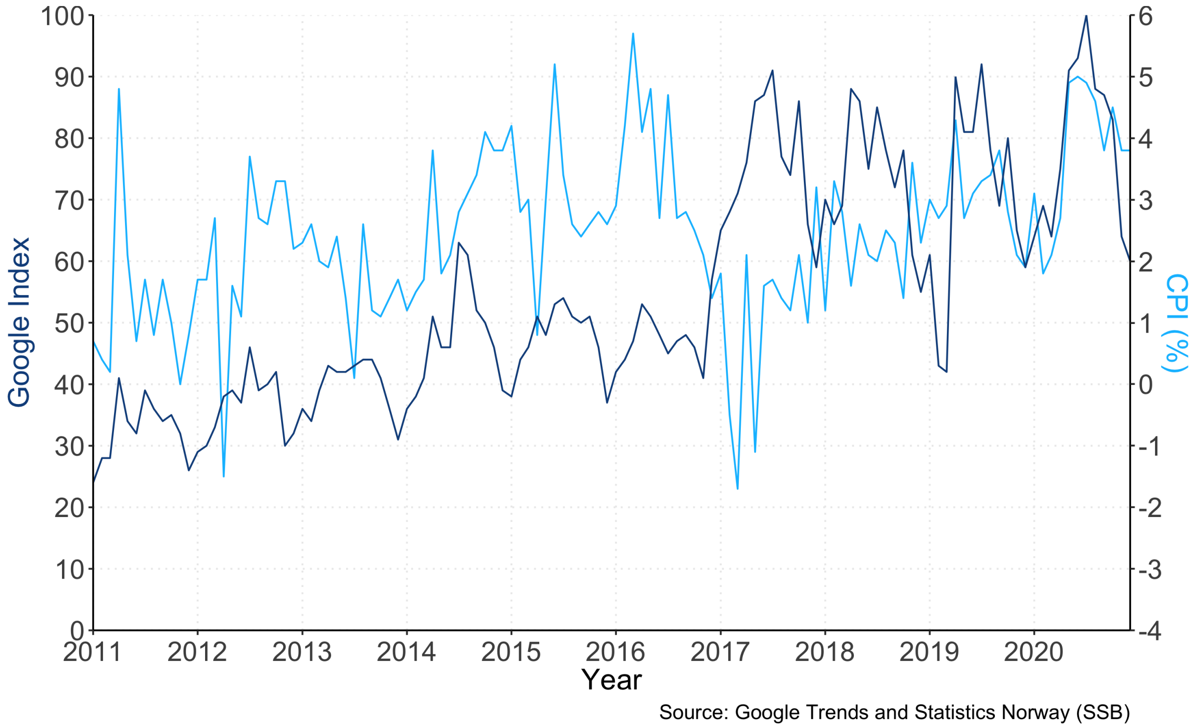
A4.6 COICOP 06 – Health

Figure A4.6: GI: Health

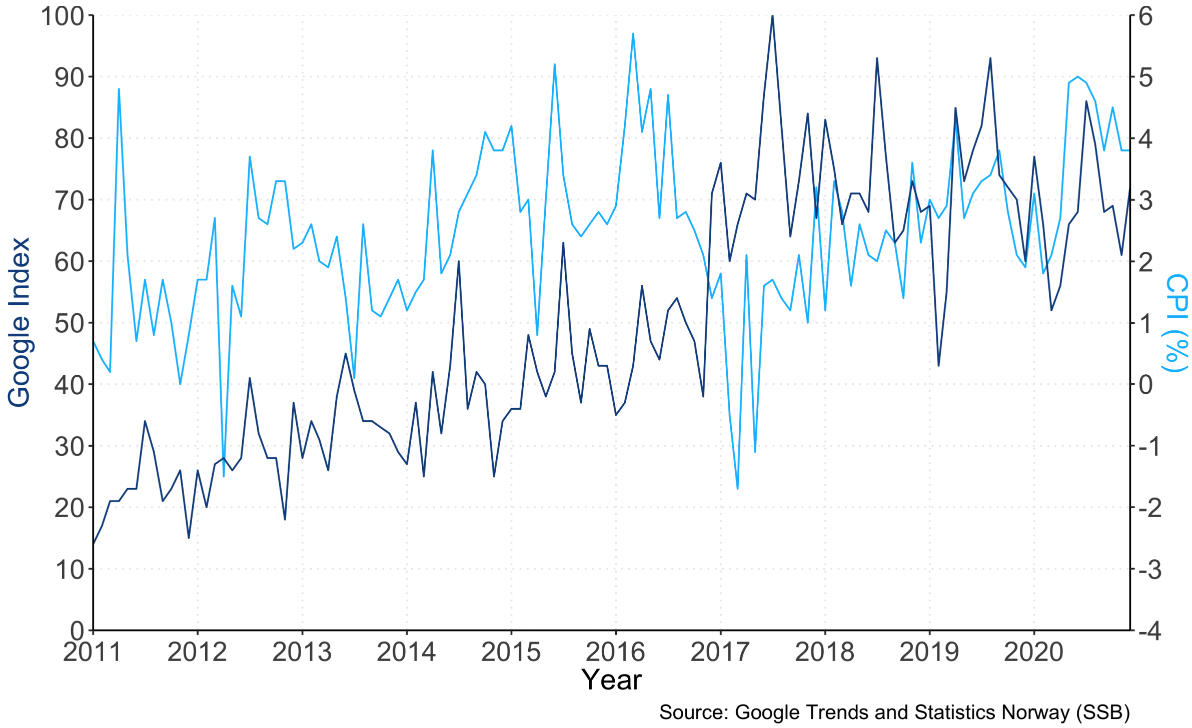


A4.7 COICOP 07 – Transport

(a) GI: Autos & vehicles

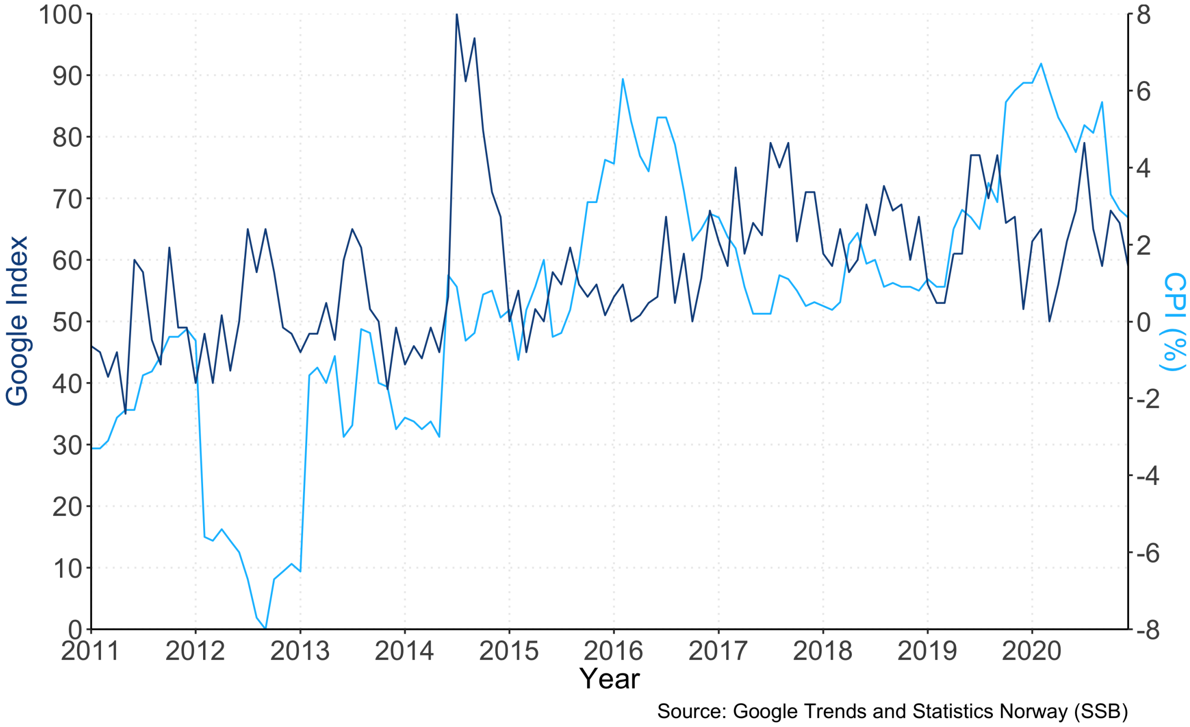


(b) GI: Transportation & logistics



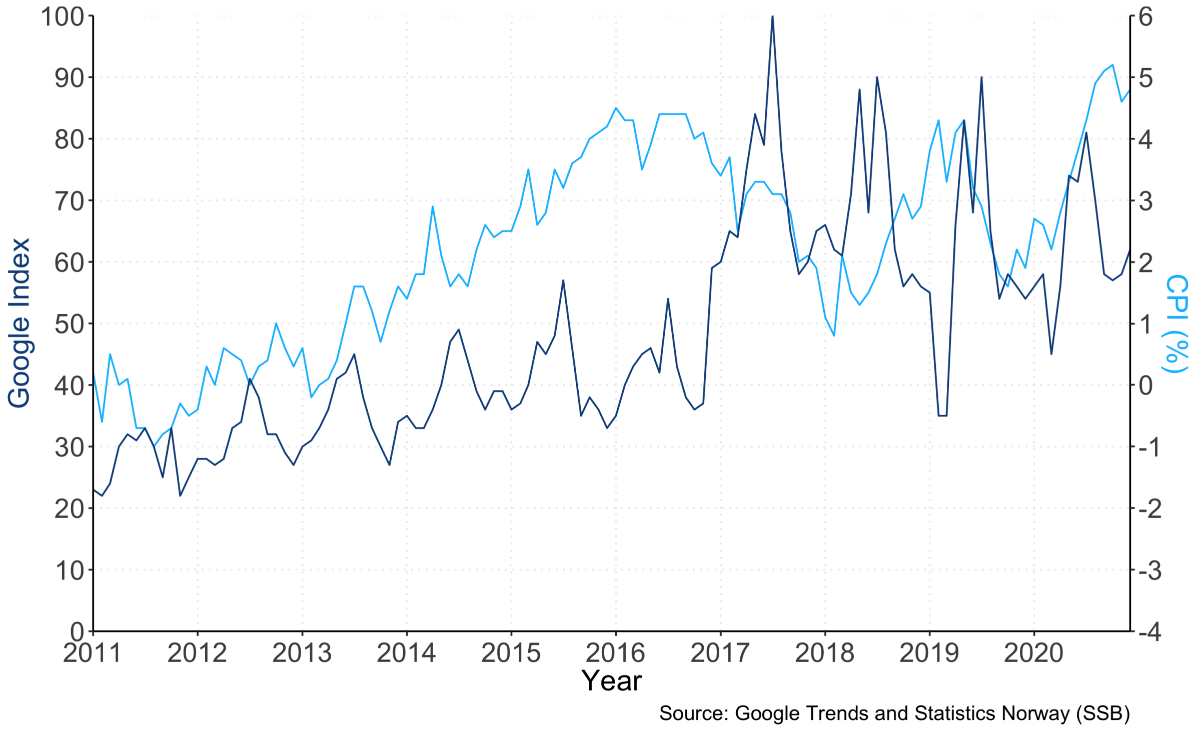
A4.8 COICOP 08 - Communications

Figure A4.8: GI: Internet & telecom



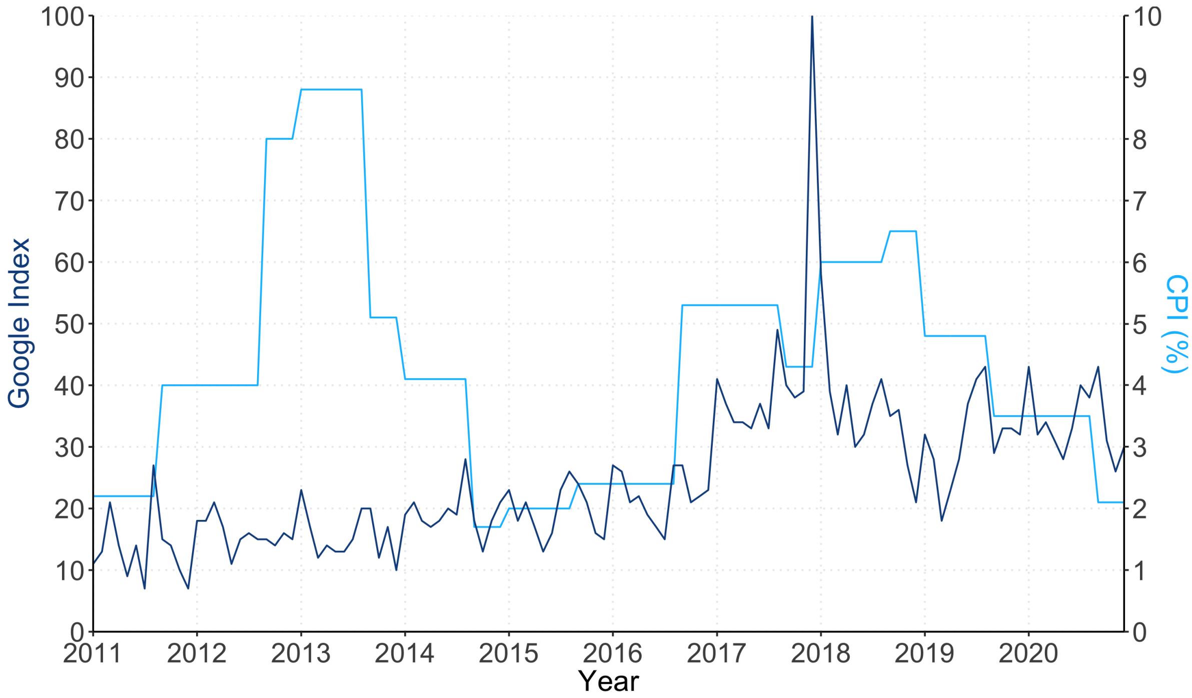
A4.9 COICOP 09 – Recreation and Culture

Figure A4.9: GI: Hobbies & leisure



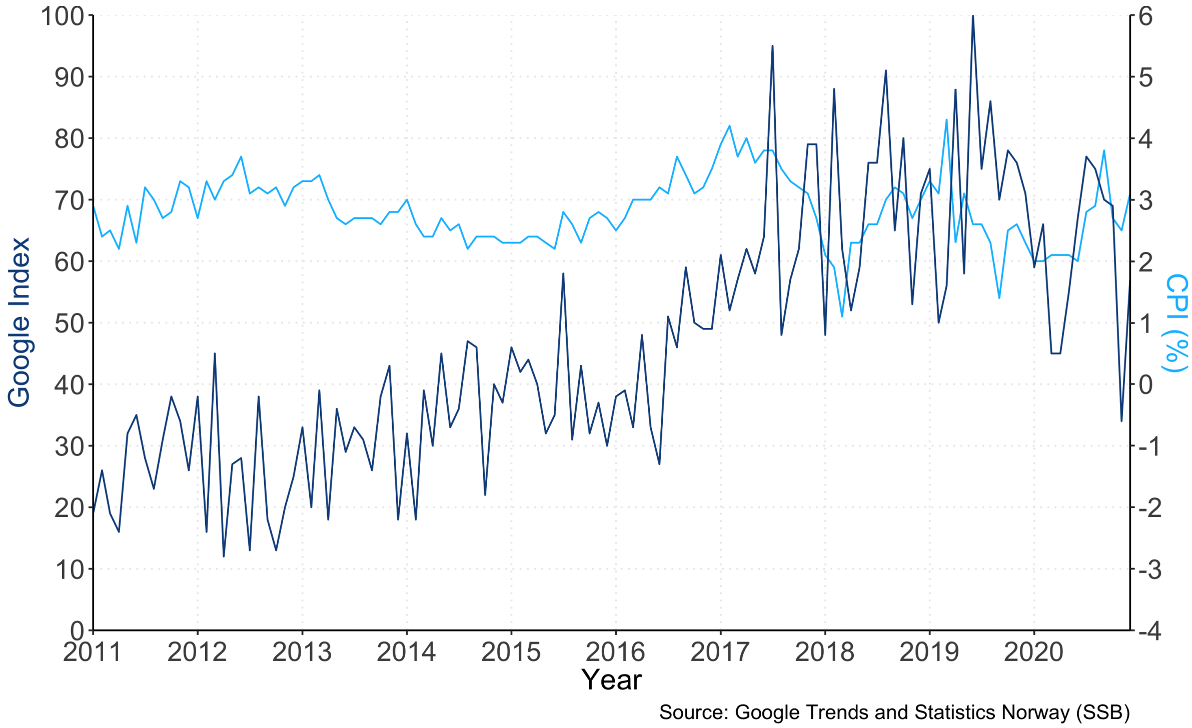
A4.10 COICOP 10 – Education

Figure A4.10: GI: Education

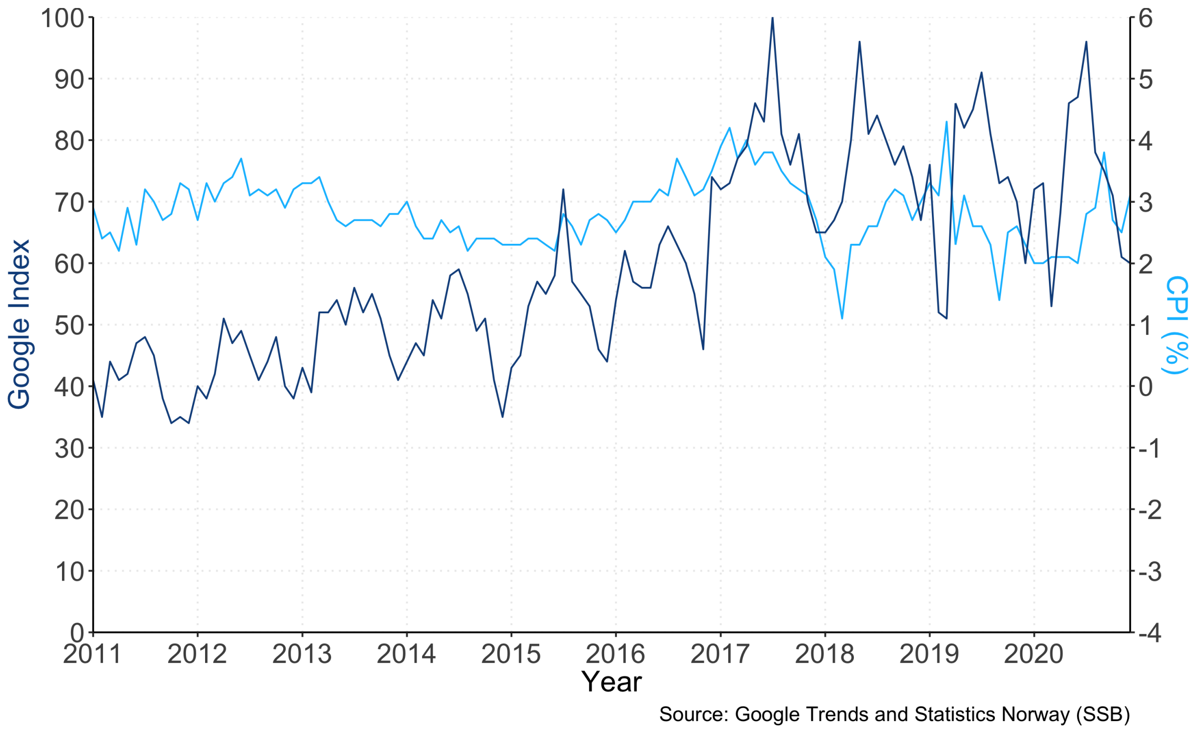


A4.11 COICOP 11 – Restaurants and Hotels

(a) GI: Restaurants

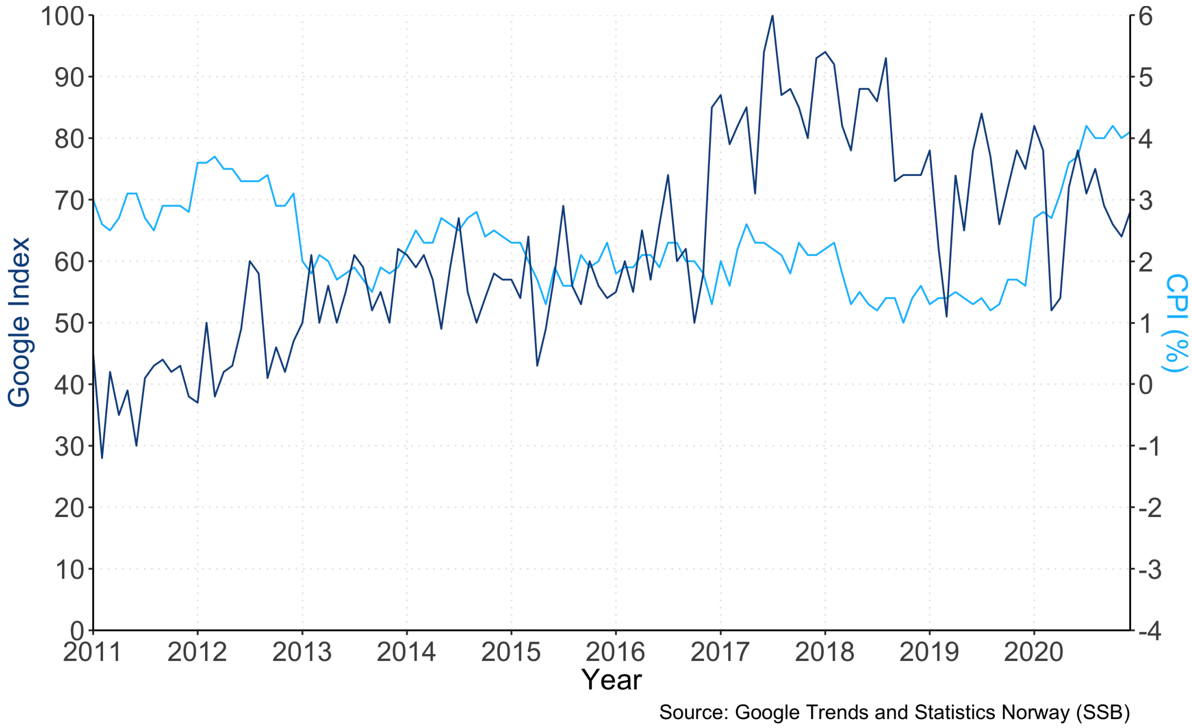


(b) GI: Hotels & accommodations

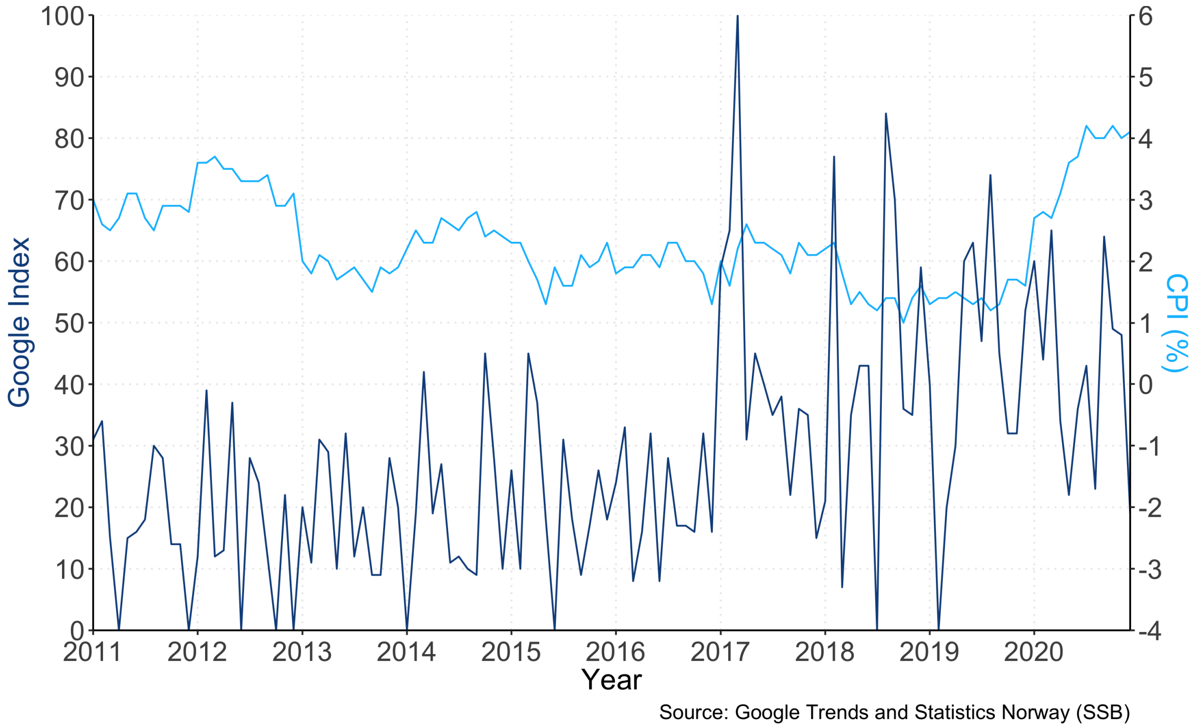


A4.12 COICOP 12 – Miscellaneous Goods and Services

(a) GI: Beauty & fitness



(b) GI: Social sercives



A5 Nowcast Benchmarks

Table A5.1: Point Nowcast 2021: COICOP 01 – Food and Non-alcoholic Beverages

	CPI Ex-Post	Extended ARIMA		AR(1)		Replicated ARIMA		Baseline ARIMA	
	COICOP 01	Model	**Nowcast**	Model	**Nowcast**	Model	**Nowcast**	Model	**Nowcast**
Jan 2021	**2.6**	(0,1,1)x(0,0,1)	**3.0**	(1,0,0)x(0,0,0)	**2.8**	(0,0,1)x(0,0,1)	**2.5**	(0,1,1)x(0,0,1)	**2.5**
Feb 2021	**2,0**	(0,1,1)x(0,0,1)	**2.4**	(1,0,0)x(0,0,0)	**2.3**	(0,0,1)x(0,0,1)	**2.4**	(0,1,1)x(0,0,1)	**2.4**
Mar 2021	**1.6**	(0,1,1)x(0,0,1)	**1.7**	(1,0,0)x(0,0,0)	**1.9**	(0,0,1)x(0,0,1)	**1.7**	(0,1,1)x(0,0,1)	**1.7**
Apr 2021	**-0.2**	(1,0,1)x(0,0,1)	**0.6**	(1,0,0)x(0,0,0)	**1.6**	(1,0,1)x(0,0,1)	**0.7**	(0,1,1)x(0,0,1)	**0.8**
May 2021	**0.0**	(1,0,1)x(0,0,1)	**0.8**	(1,0,0)x(0,0,0)	**0.3**	(1,0,1)x(0,0,1)	**0.6**	(0,1,1)x(0,0,1)	**0.7**
Jun 2021	**0.1**	(1,0,1)x(0,0,1)	**1.4**	(1,0,0)x(0,0,0)	**0.4**	(1,0,1)x(0,0,1)	**1.1**	(0,1,1)x(0,0,1)	**1.2**
Jul 2021	**-0.8**	(1,0,1)x(0,0,1)	**0.2**	(1,0,0)x(0,0,0)	**0.5**	(1,0,1)x(0,0,1)	**0.4**	(0,1,1)x(0,0,1)	**0.4**
Aug 2021	**-0.8**	(1,0,1)x(0,0,1)	**-1.3**	(1,0,0)x(0,0,0)	**-0.2**	(1,0,1)x(0,0,1)	**-1.0**	(0,1,1)x(0,0,1)	**-1.0**
Sep 2021	**-1.8**	(1,0,1)x(0,0,1)	**-1.0**	(1,0,0)x(0,0,0)	**-0.2**	(1,0,1)x(0,0,1)	**-0.8**	(0,1,1)x(0,0,1)	**-0.8**
Oct 2021	**-1.9**	(1,0,1)x(0,0,1)	**-1.6**	(1,0,0)x(0,0,0)	**-1.0**	(1,0,1)x(0,0,1)	**-1.4**	(0,1,1)x(0,0,1)	**-1.5**
Nov 2021	**-1.5**	(1,0,1)x(0,0,1)	**-1.2**	(1,0,0)x(0,0,0)	**-1.1**	(1,0,1)x(0,0,1)	**-1.1**	(1,0,1)x(0,0,1)	**-1.1**
Dec 2021	**0.4**	(1,0,1)x(0,0,1)	**-1.7**	(1,0,0)x(0,0,0)	**-0.8**	(1,0,1)x(0,0,1)	**-1.0**	(1,0,1)x(0,0,1)	**-1.0**

Table A5.2: Point Nowcast 2021: COICOP 03 – Clothing and Footwear

	CPI Ex-Post	Extended ARIMA		AR(1)		Replicated ARIMA		Baseline Model	
	COICOP 03	Model	**Nowcast**	Model	**Nowcast**	Model	**Nowcast**	Model	**Nowcast**
Jan 2021	**-0.6**	(2,1,2)x(1,0,1)	**-3.1**	(1,0,0)x(0,0,0)	**-2.3**	(2,1,2)x(1,0,1)	**-2.8**	(3,0,1)x(2,0,0)	**-4.5**
Feb 2021	**0.3**	(2,1,2)x(2,0,0)	**0.9**	(1,0,0)x(0,0,0)	**-0.6**	(2,1,2)x(2,0,0)	**0.9**	(1,0,2)x(2,0,0)	**0.9**
Mar 2021	**2.0**	(2,1,2)x(2,0,0)	**-0.6**	(1,0,0)x(0,0,0)	**0.1**	(2,1,2)x(2,0,0)	**-0.8**	(1,0,2)x(2,0,0)	**-0.3**
Apr 2021	**-2.5**	(2,1,1)x(2,0,0)	**1.6**	(1,0,0)x(0,0,0)	**1.4**	(2,1,2)x(2,0,0)	**1.6**	(1,0,0)x(2,0,0)	**2.1**
May 2021	**-1.7**	(2,1,3)x(1,0,1)	**-0.8**	(1,0,0)x(0,0,0)	**-2.1**	(2,1,3)x(1,0,1)	**-2.4**	(1,0,0)x(0,0,1)	**-1.9**
Jun 2021	**-2.9**	(2,1,3)x(1,0,1)	**-2.4**	(1,0,0)x(0,0,0)	**-1.5**	(2,1,3)x(1,0,1)	**-1.7**	(1,0,0)x(0,0,1)	**-2.5**
Jul 2021	**-1.2**	(1,0,0)x(2,0,0)	**-4.0**	(1,0,0)x(0,0,0)	**-2.4**	(1,0,0)x(2,0,0)	**-4.1**	(1,0,2)x(0,0,1)	**-4.9**
Aug 2021	**-1.4**	(1,0,2)x(0,0,1)	**-0.3**	(1,0,0)x(0,0,0)	**-1.1**	(1,0,2)x(0,0,1)	**-0.4**	(1,0,2)x(0,0,1)	**-0.2**
Sep 2021	**0.1**	(1,0,0)x(0,0,2)	**1.3**	(1,0,0)x(0,0,0)	**-1.2**	(1,0,0)x(0,0,2)	**1.3**	(1,0,2)x(0,0,1)	**0.8**
Oct 2021	**-1.2**	(1,0,2)x(1,0,1)	**-0.1**	(1,0,0)x(0,0,0)	**-0.1**	(1,0,2)x(1,0,1)	**-0.2**	(1,0,2)x(1,0,1)	**0.0**
Nov 2021	**-0.8**	(1,0,0)x(0,0,2)	**-0.6**	(1,0,0)x(0,0,0)	**-1.1**	(1,0,0)x(0,0,2)	**-0.6**	(1,0,2)x(1,0,1)	**-0.4**
Dec 2021	**-1.2**	(1,0,0)x(0,0,2)	**-1.2**	(1,0,0)x(0,0,0)	**-0.8**	(1,0,0)x(0,0,2)	**-1.1**	(1,0,2)x(1,0,1)	**-0.9**

Table A5.3: Point Nowcast 2021: COICOP 04 – Housing, Water, Electricity, Gas and Other Fuels

	CPI Ex-Post	Extended ARIMA		AR(1)		Replicated ARIMA		Baseline ARIMA	
	COICOP 04	Model	**Nowcast**	Model	**Nowcast**	Model	**Nowcast**	Model	**Nowcast**
Jan 2021	**1.0**	(1,1,0)x(0,0,1)	**1.1**	(1,0,0)x(0,0,0)	**1.2**	(1,1,0)x(0,0,1)	**1.2**	(1,1,0)x(0,0,1)*	**1.2**
Feb 2021	**0.8**	(1,1,0)x(0,0,1)	**1.1**	(1,0,0)x(0,0,0)	**1.0**	(1,1,0)x(0,0,1)	**1.1**	(1,1,0)x(0,0,2)*	**1.0**
Mar 2021	**0.9**	(0,1,0)x(0,0,1)	**0.8**	(1,0,0)x(0,0,0)	**0.8**	(0,1,0)x(0,0,1)	**0.9**	(0,1,0)x(0,0,2)*	**0.8**
Apr 2021	**0.9**	(0,1,0)x(0,0,1)	**0.9**	(1,0,0)x(0,0,0)	**0.9**	(0,1,0)x(0,0,1)	**0.9**	(0,1,0)x(0,0,2)*	**0.9**
May 2021	**1.3**	(0,1,0)x(0,0,1)	**1.0**	(1,0,0)x(0,0,0)	**0.8**	(0,1,0)x(0,0,1)	**1.0**	(0,1,0)x(0,0,2)*	**1.0**
Jun 2021	**1.3**	(1,1,0)x(0,0,1)	**1.2**	(1,0,0)x(0,0,0)	**1.3**	(1,1,0)x(0,0,1)	**1.2**	(1,1,0)x(0,0,1)*	**1.2**
Jul 2021	**1.6**	(1,1,0)x(0,0,1)	**1-3**	(1,0,0)x(0,0,0)	**1.3**	(1,1,0)x(0,0,1)	**1.4**	(1,1,0)x(0,0,1)*	**1.4**
Aug 2021	**1.3**	(1,1,0)x(0,0,1)	**1.4**	(1,0,0)x(0,0,0)	**1.6**	(1,1,0)x(0,0,1)	**1.4**	(1,1,0)x(0,0,1)*	**1.4**
Sep 2021	**1.4**	(1,1,0)x(0,0,1)	**1.3**	(1,0,0)x(0,0,0)	**1.3**	(1,1,0)x(0,0,1)	**1.3**	(1,1,0)x(0,0,1)*	**1.3**
Oct 2021	**1.6**	(1,1,0)x(0,0,1)	**1.5**	(1,0,0)x(0,0,0)	**1.3**	(1,1,0)x(0,0,1)	**1.5**	(1,1,0)x(0,0,1)*	**1.5**
Nov 2021	**1.6**	(1,1,0)x(0,0,1)	**1.6**	(1,0,0)x(0,0,0)	**1.6**	(1,1,0)x(0,0,1)	**1.6**	(1,1,0)x(0,0,1)*	**1.6**
Dec 2021	**1.5**	(1,1,0)x(0,0,1)	**1.5**	(1,0,0)x(0,0,0)	**1.6**	(1,1,0)x(0,0,1)	**1.5**	(1,1,0)x(0,0,1)*	**1.5**

Table A5.4: Point Nowcast 2021: COICOP 05 – Furnishings, Household, and Routine Maintenance

	CPI Ex-Post	Extended ARIMA		AR(1)		Replicated ARIMA		Baseline ARIMA	
	COICOP 05	Model	**Nowcast**	Model	**Nowcast**	Model	**Nowcast**	Model	**Nowcast**
Jan 2021	**6.2**	(0,1,2)x(1,0,1)	**5.8**	(1,0,0)x(0,0,0)	**7.5**	(0,1,2)x(1,0,1)	**5.6**	(0,1,2)x(1,0,1)*	**5.7**
Feb 2021	**7.0**	(0,1,2)x(1,0,1)	**7.5**	(1,0,0)x(0,0,0)	**5.9**	(0,1,2)x(1,0,1)	**7.2**	(0,1,2)x(1,0,1)*	**7.4**
Mar 2021	**7.6**	(0,1,2)x(1,0,1)	**7.2**	(1,0,0)x(0,0,0)	**6.6**	(0,1,2)x(1,0,1)	**7.0**	(0,1,2)x(1,0,1)*	**7.2**
Apr 2021	**5.6**	(0,1,2)x(1,0,1)	**6.4**	(1,0,0)x(0,0,0)	**7.2**	(0,1,2)x(1,0,1)	**5.9**	(0,1,2)x(1,0,1)*	**6.0**
May 2021	**2.8**	(0,1,2)x(1,0,1)	**6.6**	(1,0,0)x(0,0,0)	**5.3**	(0,1,2)x(1,0,1)	**6.2**	(0,1,2)x(1,0,1)*	**6.4**
Jun 2021	**3.6**	(1,1,2)x(0,0,1)	**3.1**	(1,0,0)x(0,0,0)	**2.7**	(1,1,2)x(0,0,1)	**3.2**	(0,1,2)x(0,0,1)	**3.3**
Jul 2021	**3.3**	(1,1,2)x(0,0,1)	**2.4**	(1,0,0)x(0,0,0)	**3.5**	(1,1,2)x(0,0,1)	**2.7**	(0,1,2)x(0,0,1)	**2.2**
Aug 2021	**0.3**	(1,1,2)x(0,0,1)	**1.6**	(1,0,0)x(0,0,0)	**3.2**	(1,1,2)x(0,0,1)	**1.8**	(1,1,2)x(0,0,1)	**1.8**
Sep 2021	**2.7**	(1,1,2)x(0,0,1)	**1.1**	(1,0,0)x(0,0,0)	**0.5**	(1,1,2)x(0,0,1)	**1.3**	(1,1,2)x(0,0,1)	**1.3**
Oct 2021	**1.1**	(3,1,1)x(0,0,1)	**0.9**	(1,0,0)x(0,0,0)	**2.7**	(3,1,1)x(0,0,1)	**0.9**	(3,1,1)x(0,0,1)	**0.9**
Nov 2021	**1.8**	(3,1,1)x(0,0,1)	**2.4**	(1,0,0)x(0,0,0)	**1.2**	(3,1,1)x(0,0,1)	**2.4**	(3,1,1)x(0,0,1)	**2.4**
Dec 2021	**2.8**	(3,1,1)x(0,0,1)	**2.0**	(1,0,0)x(0,0,0)	**1.8**	(3,1,1)x(0,0,1)	**2.1**	(3,1,1)x(0,0,1)	**2.1**

Table A5.5: Point Nowcast 2021: COICOP 06 – Health

	CPI Ex-Post	Extended ARIMA		AR(1)		Replicated ARIMA		Baseline ARIMA	
	COICOP 06	Model	**Nowcast**	Model	**Nowcast**	Model	**Nowcast**	Model	**Nowcast**
Jan 2021	**2.0**	(1,0,1)x(0,0,1)	**1.3**	(1,0,0)x(0,0,0)	**1.8**	(1,0,1)x(0,0,1)	**1.6**	(1,0,3)x(0,0,1)	**1.7**
Feb 2021	**2.7**	(1,0,1)x(2,0,0)	**1.9**	(1,0,0)x(0,0,0)	**2.1**	(1,0,1)x(2,0,0)	**1.9**	(1,0,3)x(0,0,1)	**1.8**
Mar 2021	**3.3**	(1,0,1)x(2,0,0)	**2.5**	(1,0,0)x(0,0,0)	**2.6**	(1,0,1)x(2,0,0)	**2.5**	(1,0,3)x(0,0,1)	**2.5**
Apr 2021	**3.3**	(1,0,1)x(2,0,0)	**3.1**	(1,0,0)x(0,0,0)	**3.1**	(1,0,1)x(2,0,0)	**3.1**	(1,0,3)x(0,0,1)	**3.0**
May 2021	**3.7**	(1,0,0)x(2,0,0)	**3.3**	(1,0,0)x(0,0,0)	**3.1**	(1,0,1)x(2,0,0)	**3.3**	(4,0,3)x(0,0,1)	**3.5**
Jun 2021	**3.3**	(1,0,1)x(2,0,0)	**3.4**	(1,0,0)x(0,0,0)	**3.5**	(1,0,1)x(2,0,0)	**3.4**	(4,0,3)x(0,0,1)	**3.6**
Jul 2021	**3.2**	(1,0,1)x(2,0,0)	**3.3**	(1,0,0)x(0,0,0)	**3.1**	(1,0,1)x(2,0,0)	**3.3**	(1,0,3)x(0,0,1)	**3.4**
Aug 2021	**2.5**	(1,0,1)x(2,0,0)	**3.6**	(1,0,0)x(0,0,0)	**3.1**	(1,0,1)x(2,0,0)	**3.6**	(1,0,3)x(0,0,1)	**3.2**
Sep 2021	**3.0**	(1,0,0)x(2,0,0)	**2.7**	(1,0,0)x(0,0,0)	**2.5**	(1,0,0)x(2,0,0)	**2.7**	(4,0,3)x(0,0,1)	**2.6**
Oct 2021	**2.6**	(1,0,0)x(2,0,0)	**2.7**	(1,0,0)x(0,0,0)	**2.9**	(1,0,0)x(2,0,0)	**2.7**	(4,0,3)x(0,0,1)	**2.4**
Nov 2021	**2.9**	(1,0,0)x(2,0,0)	**3.0**	(1,0,0)x(0,0,0)	**2.6**	(1,0,0)x(2,0,0)	**3.0**	(1,0,3)x(0,0,1)	**2.7**
Dec 2021	**3.2**	(1,0,0)x(2,0,0)	**2.8**	(1,0,0)x(0,0,0)	**2.8**	(1,0,0)x(2,0,0)	**2.8**	(1,0,3)x(0,0,1)	**3.0**

Table A5.6: Point Nowcast 2021: COICOP 07 - Transport

	CPI Ex-Post	Extended ARIMA		AR(1)		Replicated ARIMA		Baseline ARIMA	
	COICOP 07	Model	**Nowcast**	Model	**Nowcast**	Model	**Nowcast**	Model	**Nowcast**
Jan 2021	**3.2**	(1,0,1)x(0,0,1)	**3.4**	(1,0,0)x(0,0,0)	**3.1**	(1,0,1)x(0,0,1)	**3.2**	(0,1,1)x(0,0,1)	**3.6**
Feb 2021	**3.7**	(1,0,1)x(0,0,1)	**4.2**	(1,0,0)x(0,0,0)	**2.8**	(1,0,1)x(0,0,1)	**3.9**	(0,1,1)x(0,0,1)	**4.2**
Mar 2021	**3.1**	(1,0,1)x(0,0,1)	**3.8**	(1,0,0)x(0,0,0)	**3.1**	(1,0,1)x(0,0,1)	**3.6**	(0,1,1)x(0,0,1)	**3.8**
Apr 2021	**1.4**	(1,0,1)x(0,0,1)	**3.2**	(1,0,0)x(0,0,0)	**2.8**	(1,0,1)x(0,0,1)	**2.2**	(0,1,1)x(0,0,1)	**2.5**
May 2021	**0.3**	(1,0,1)x(0,0,1)	**1.8**	(1,0,0)x(0,0,0)	**1.9**	(1,0,1)x(0,0,1)	**1.1**	(0,1,1)x(0,0,1)	**1.3**
Jun 2021	**-0.7**	(1,0,1)x(0,0,1)	**0.4**	(1,0,0)x(0,0,0)	**1.3**	(1,0,1)x(0,0,1)	**0.2**	(0,1,1)x(0,0,1)	**0.3**
Jul 2021	**-1.7**	(1,0,1)x(0,0,1)	**-0.2**	(1,0,0)x(0,0,0)	**0.8**	(1,0,1)x(0,0,1)	**-0.5**	(0,1,1)x(0,0,1)	**-0.4**
Aug 2021	**-0.6**	(1,0,1)x(0,0,1)	**-0.5**	(1,0,0)x(0,0,0)	**0.1**	(1,0,1)x(0,0,1)	**-0.6**	(1,0,1)x(0,0,1)	**-0.6**
Sep 2021	**-0.6**	(1,0,1)x(0,0,1)	**0.0**	(1,0,0)x(0,0,0)	**0.6**	(1,0,1)x(0,0,1)	**0.1**	(1,0,1)x(0,0,1)	**0.1**
Oct 2021	**-0.5**	(1,0,1)x(0,0,1)	**0.1**	(1,0,0)x(0,0,0)	**0.6**	(1,0,1)x(0,0,1)	**-0.1**	(1,0,1)x(0,0,1)	**-0.1**
Nov 2021	**1.2**	(1,0,1)x(0,0,1)	**-0.2**	(1,0,0)x(0,0,0)	**0.6**	(1,0,1)x(0,0,1)	**0.2**	(1,0,1)x(0,0,1)	**0.2**
Dec 2021	**2.0**	(1,0,1)x(0,0,1)	**0.6**	(1,0,0)x(0,0,0)	**1.6**	(1,0,1)x(0,0,1)	**0.9**	(1,0,1)x(0,0,1)	**0.9**

A6 Descriptive Statistics: Out-of-Sample Nowcasts

Table A6.1: Descriptive Statistics: COICOP 01, 03 and 04

	CPI ex-post	Extended ARIMA	Replicated ARIMA	Baseline ARIMA	AR(1)
COICOP 01 - Food and Non-alcoholic Beverages					
Mean Point Estimates	-0.02	0.27	0.34	0.35	0.54
SD Point Estimate	1.47	2.64	1.90	1.99	1.72
RMSE		0.90	0.77	0.79	0.94
Correct Trends		0.82	0.82	0.82	0.73
Turning Points		0.60	0.50	0.50	0.30
COICOP 03 - Clothing and Footwear					
Mean Point Estimates	-0.93	-0.77	-0.86	-0.99	-0.98
SD Point Estimate	0.95	2.96	2.87	4.53	1.16
RMSE		1.91	1.92	2.23	1.57
Correct Trends		0.45	0.36	0.45	0.18
Turning Points		0.10	0.10	0.10	0.10
COICOP 04 - Housing, Water, Electricity, Gas and Other Fuels					
Mean Point Estimates	1.27	1.24	1.26	1.23	1.26
SD Point Estimate	0.31	0.07	0.06	0.06	0.09
RMSE		0.16	0.17	0.17	0.22
Correct Trends		0.73	0.73	0.82	0.55
Turning Points		0.50	0.50	0.60	0.20

Table A6.2: Descriptive Statistics: COICOP 05, 06 and 07

	CPI ex-post	Extended ARIMA	Replicated ARIMA	Baseline ARIMA	AR(1)
COICOP 05 - Furnishings, Household, and Routine Maintenance					
Mean Point Estimates	3.73	3.92	3.84	3.88	4.01
SD Point Estimate	2.35	6.53	5.46	6.02	5.81
RMSE		1.36	1.22	1.29	1.60
Correct Trends		0.55	0.55	0.55	0.36
Turning Points		0.20	0.20	0.20	0.00
COICOP 06 - Health					
Mean Point Estimates	2.98	2.80	2.82	2.78	2.76
SD Point Estimate	0.46	0.44	0.37	0.40	1.22
RMSE		2.94	2.93	2.89	2.79
Correct Trends		0.64	0.64	0.82	0.64
Turning Points		0.40	0.40	0.60	0.40
COICOP 07 - Transport					
Mean Point Estimates	0.90	1.41	1.19	1.31	1.61
SD Point Estimate	1.80	3.11	2.66	3.05	1.22
RMSE		1.14	0.74	0.83	1.26
Correct Trends		0.82	0.82	0.82	0.64
Turning Points		0.60	0.60	0.60	0.40