NOWCASTING GDP GROWTH USING GOOGLE TRENDS

Jyväskylä University
School of Business and Economics

Master's Thesis

2019

Author: Joni Heikkinen
Subject: Economics
Supervisor: Kari Heimonen/Petteri Juvonen

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

ABSTRACT

Author Joni Heikkinen	
Title Nowcasting GDP growth using Google Trends	
Subject Economics	Type of work Master's Thesis
Date 10/21/2019	Number of pages 80+19

Abstract

This master's thesis examines Google Trends ability to nowcast Germany and Finland's economic growth, i.e. gross domestic product (GDP). Nowcasting aims to forecast the current economic situation. Google Trends data reflects the popularity of different Google searches. Early studies found that Google Trends can generate accurate forecasts for various economic variables, many of which are related to GDP. In this regard, Götz and Knetsch (2019) used Google Trends data to nowcast Germany's GDP. GDP is an important economic variable that is published quarterly and has a significant publication delay. However, economic changes can occur quickly and suddenly. Therefore, it is important to obtain up-to-date information about the current economic situation.

In addition to Google Trends data, this study uses Germany and Finland's consumer confidence data as a benchmark. This master's thesis follows Götz and Knetsch' (2019) study closely and selects similar initial search categories. A large number of initial search categories causes the problem of high dimensionality. This thesis solves the problem by using both dimension reduction and variable selection methods. The master's thesis answers to the research topic by creating a nowcasting exercise that attempts to simulate a real-life nowcasting situation. Exercise will include multiple nowcasting models, which this thesis examines with their root mean square errors and figures.

According to the results of this master's thesis, the most accurate model for a broad Google category model was the "News" model. The models were also examined in subcategory levels. The "Banking" model was the most precise subcategory model in Finland. In Germany, however, the "Vehicle Shows" category was the most accurate subcategory. Overall, Google models perform significantly better in Germany than in Finland, where consumer confidence data provided very accurate predictions. Moreover, the thesis evaluated leading models with a leave-one-out cross-validation method, which confirmed previous results, i.e. in both countries, the consumer confidence was the leading model. Furthermore, Donadelli (2015) found that Google searches had a relationship with policy-related uncertainty. This study did not find a similar relationship.

Key words
Nowcasting, forecasting, GDP, economic growth, Google Trends

Place of storage
Jyväskylä University Library

TIIVISTELMÄ

Tekijä Joni Heikkinen	
Työn nimi Nowcasting GDP growth using Google Trends	
Oppiaine Taloustiede	Työn laji Pro gradu -tutkielma
Päivämäärä 21.10.2019	Sivumäärä 80+19

Tiivistelmä

Tässä Pro gradu -tutkielmassa tutkitaan Google Trends -aineiston kykyä nowcasting ennustaa Saksan ja Suomen talouskasvua eli bruttokansantuotetta (BKT). Nowcasting pyrkii ennustamaan nykyistä taloudellista tilannetta. Google Trends -aineisto kuvaa taas erilaisten Google-hakujen suosiota. Varhaisissa tutkimuksissa havaittiin, että Google Trends -data voi tuottaa tarkkoja ennusteita monille taloudellisille muuttujille, joista monet liittyvät BKT:hen. Tähän liittyen, Götz & Knetsch (2019) käyttivät Google Trends -dataa Saksan BKT:n nowcasting ennustamiseen. BKT on tärkeä taloudellinen muuttuja, jolla on huomattava julkistamisviive. Taloudelliset muutokset voivat kuitenkin tapahtua nopeasti ja yllättävästi, ja siksi on tärkeää saada ajankohtaisempaa tietoa talouden tilasta.

Google-hakudatan lisäksi tässä Pro gradu -tutkielmassa käytetään vertailukohteena kuluttajien luottamus -aineistoa. Tämä tutkielma seuraa Götz ja Knetsch (2019) tutkimusta ja valitsee samat alustavat hakukategoriat. Hakukategorioiden suuri lukumäärä aiheuttaa korkeaulotteisen aineiston ongelman. Tutkielma ratkaisee korkeaulotteisen aineiston ongelman käyttämällä sekä ulottuvuuden supistamis- että muuttujan valikointi -menetelmiä. Tutkimuskysymykseen vastatakseen Pro gradu -tutkielma luo nowcasting-ennusteharjoituksen, joka pyrkii simuloimaan todellista ennustetilannetta. Ennusteharjoituksessa käytettiin lukuisia ennustemalleja, joita vertailtiin niiden keskineliövirheen neliöjuurilla ja kuviolla.

Tämän Pro gradu -tutkielman tulosten mukaan tarkin laajan Google-hakukategoriamalli oli ”Uutiset”-malli. Suomen tarkimmaksi alakategoriaksi paljastui ”Pankkitoiminta”-alakategoria. Saksassa taas ”Automessut”-kategoria oli tarkin alakategoria. Google-mallit toimivat paremmin Saksassa kuin Suomessa, jossa kuluttajien luottamus -aineisto tuotti johdonmukaisesti tarkempia ennusteita. Parhaimpia malleja arvioitiin myös ristiinvalidoinnilla, joka vahvisti aikaisemmat tulokset, ts. molemmissa maissa kuluttajien luottamus oli tarkin nowcasting-malli. Donadelli (2015) havaitsi, että Google-hauilla olisi yhteys politiikkaan liittyvään taloudelliseen epävarmuuteen. Tämä Pro gradu -tutkielma ei kuitenkaan havainnut yhtä vahvaa yhteyttä.

Asiasanat
Nowcasting, ennustaminen, BKT, talouskasvu, Google Trends

Säilytyspaikka
Jyväskylän yliopiston kirjasto

CONTENTS

LIST OF TABLES AND FIGURES

1 INTRODUCTION

Government's statistics agencies publish economic statistics with a significant delay. For example, Statistics Finland publishes Finland's gross domestic product quarterly, at best (Statistics Finland, 2019b). This causes two-month publication delay after the end of the quarter. However, changes in economic conditions can happen swiftly and suddenly. Therefore, it is in policymakers and central banks' interests to have more timely statistics on the current economic situation.

Nowcasting attempts to forecast macroeconomic variables even months before their initial publishing (Koop & Onorante, 2013). Nowcasting is not trying to forecast the future; instead, it is trying to predict the present economic situation (Choi & Varian, 2012). One can also consider nowcasting models as providing predictions about the very near past and future (Bańbura, Giannone, Modugno, & Reichlin, 2013, 196). To create these nowcasting forecasts, nowcasting models demand more timely data sources, i.e. monthly, weekly or daily data.

One of the timeliest sources is unstructured data called "*Big data*", which is generated among other things from extensive internet usage. Nowadays more and more people have a device, for example, a mobile phone, which they can use to access the internet. In 2004, 1.7 billion people had a cellular subscription, and in 2016, subscriptions had increased to over 7.6 billion (World Bank, 2019b). These developments in mobile devices have increased the amount of the world's population, which have access to the internet. World Bank estimates that in 2016, 46 % of the world's population had access to the internet when in 2004; the estimate was just 14 % (World Bank, 2019a).

Moreover, internet access has improved substantially also in European countries. In 2007, 55 % of European household had access to the internet. Later in 2016, the share had increased to 85 %. The EU-Member countries with the most substantial internet access were Netherlands and Luxembourg, with 97 % of households having access to the internet (Eurostat, 2018). These statistics paint the picture that the internet has become a regular part of our daily lives.

One of the popular uses of the internet is searching for information, for example, searching for appropriate housing, booking hotels, buying products and even for dating. Eurostat survey estimates that 80 % of European internet users aged 16– 74 have used it for searching for information (Eurostat, 2018). Therefore, one of the most notable big data sources generated by Google searches. Google search engine, developed by Google LLC, is the most used internet search engine in the world (Statista, 2018).

In 2010, Google LLC revealed that the Google search engine was proving over a billion searches per day (Google, 2010). It is safe to assume that these searches have only increased to this day; consequently, this has created one of the world's largest databases. Fortunately, Google LLC has made this vast database publicly available. Since 2004, Google LLC has published their search data on its *Google Trends* website[1].

[1] https://trends.google.com

Google Trends data are available in a quite extensive form since Google publishes search data weekly and daily in real-time. Additionally, the user can specify the data by country and in some cases, even in the municipality level. (Google, 2019a; Choi and Varian, 2012.)

Econometric literature has extensively studied Google Trends data's potential abilities, and the early results have been quite promising. Google Trends data has been able to nowcast such macroeconomic variables as unemployment and consumption. Furthermore, it has also been used to nowcast travelling, car sales and even the financial markets.

However, studies using Google Trends data for nowcasting countries economic growth, i.e. gross domestic product (GDP), are relatively rare. Even though GDP is one of the most used a macroeconomic variable in economic research. Therefore, it is of great importance to predict the current and near-future GDP. Earlier studies have also found Google Trends data improving model's prediction accuracy in different sectors of the economy, for example, travelling and automobiles. GDP includes the output of these industries; hence, Google Trends could also be used to nowcast the country's GDP. Moreover, GDP is a very influential economic variable; thus, even a small improvement nowcasting accuracy could lead to considerable economic benefits.

It also seems that nowcasting literature has finally started to study the use of Google Trends to predict GDP. In 2019, Götz and Knetsch (2019) published the first known study in which, they investigated Google Trend data's ability to nowcast German GDP using bridge equation models. According to their results, Google Trend variables provide additional information for long and mid-term GDP forecasts. (Götz & Knetsch, 2019, 53–54.)

Another study by Ferrara and Simoni (2019) found that Google Trends variables provide useful information for the first four weeks of the forecasting period. On the contrary to previous results, this suggests that Google Trends data is especially valuable for the short-run forecasts. These results present quite picture mixed about the Google Trends data's abilities to nowcast the country's GDP growth. Therefore, it would be interesting to shed new light into this discussion and study if Google Trends data is any good in nowcasting a country's GDP growth.

For a more comprehensive analysis, this master's thesis uses two different countries to examine Google Trends nowcasting ability. The first country is Germany, which is akin to earlier studies, i.e. Götz & Knetsch (2019), Ferrara, and Simoni (2019). In this thesis, Germany also represents a large open economy that has a wide arrange of different industries.

The second country where this thesis examines Google Trends is Finland, which represents a small and even more open economy. In addition, Finland is a relevant country to study this matter because it has one of the highest internet access in Europe (Eurostat, 2018). It also estimated that Finnish residents are using the internet even more often (Statistics Finland, 2016). This extensive internet use should produce interesting search data, which in turn, this thesis applies for its research.

More strictly, this master's thesis studies whether Google Trends data provides sound nowcasting forecasts for both Germany and Finland's economic growth, i.e. gross domestic product (GDP). This thesis is not trying to find a causal relationship between Google Trends data and GDP. Instead, the focus is to examine whether Google Trends data could produce additional information concerning the current economic conditions. This thesis also compares Google Trends data to consumer survey data, which enables a more precise and robust examination.

Explicitly stated, this study assumes that Google Trends data is a proxy for peoples' interest in durable goods. For example, the more searches there are for Autos & Vehicles the consumers are signalling higher willingness to buy new cars and trucks. This increased consumption leads to increases in economic growth, as the automobile industry is a part of the GDP.

The master's thesis has the following structure. The first chapter provides an introduction and motivation to the research theme. The second chapter is a literature review of previous studies regarding nowcasting and studies that have examined Google Trends data. After that, the third chapter describes the thesis data series that include official GDP statistics, consumer confidence statistics and Google Trends data.

Forth chapter illustrates the research methods that this thesis used. Google Trends data was highly dimensional, i.e. it included a large number of variables. Therefore, this thesis used dimension reduction and shrinkage methods. This thesis also conducted nowcasting exercise to study Google Trends data's forecasting abilities. Nowcasting exercise was estimated in pseudo-out-of-sample to simulate real nowcast situation. The results of these estimates are in the subsequent chapter five. The final chapter concludes this thesis and presents some suggestions for further studies.

2 LITERATURE REVIEW

This literature review intends to provide a concise overview of nowcasting, Google Trends literature and present new studies regarding Google Trends use in nowcasting GDP. The literature review begins with an introduction to the nowcasting theorem and Google Trends data's potential benefit to it. The subsequent section provides a brief examination concerning the Google Trends data studies and their progression. Finally, this literature review discusses new studies that have inspected Google Trends abilities to nowcast GDP growth.

2.1 Nowcasting

The nowcasting literature started with simple models nowcasting quarterly variables using monthly data series. Trehan (1989) used a bridge equation model to nowcast the United States gross national product (GNP). In other words, Trehan (2019, 42–43) predicted first the selected monthly variables, for example, industrial production and then the quarterly GNP. Rünstler & Sedillot (2003) applied similar types of bridge models for the Euro area's countries. More strictly speaking, they attempted to nowcast Euro areas quarterly GDP growth using multiple monthly data series (Rünstler & Sedillot, 2003).

However, Evans' (2005) study was a significant turning point for the nowcasting literature because it was one of the first to estimate both the United States short-term GDP growth and level. More importantly, Evans' (2005) statistical model considered that information regarding GDP becomes available at different periods. Since information sets are available in different time periods, there are missing observations in some periods, which leads to an unbalanced data set. The nowcasting literature calls this the *ragged-edge database* problem (Giannone, Reichlin & Small, 2008).

Rünstler & Sedillot (2003) had previously modelled missing observations through different time series models. But Evans (2005) had a new innovative solution. To solve the issue, Evans (2005) applied the Kalman filter, which provides estimates for the missing observations. With the Kalman filter, Evans (2005) estimated the model with 19 different macroeconomic variables regarding GDP growth. However, in practice, short-term forecasters use even larger information sets and variables. For example, the Bank of Finland uses 48 different variables for nowcasting Finland's GDP (Itkonen & Juvonen, 2017).

With this number of variables, Evans' (2005) model could lead to overfitting, i.e. variables would start to weaken the model's forecasts. A large number of model variables would also increase Evans' (2005) model's estimation uncertainty. Consequently, Giannone, Reichlin & Small (2008) refined Evans' (2005) model by presenting a new dynamic factor model (DFM). This new DFM based on an earlier study by Doz, Giannone & Reichlin (2006).

In any case, Giannone, Reichlin & Small (2008) DFM allows for an even more significant number of information sets, i.e. variables. For this reason, the dynamic factor model has been quite popular among central banks, e.g. the Federal Reserve Bank of New York.

Giannone et al. (2008) dynamic factor model has a two-part structure. First, the multiple information sets are reduced to a common factor by a principal component analysis (PCA). According to Giannone et al. (2008), PCA provides adequate approximations about the optimal model, and it does not lead to over-parametrization. In other words, PCA uses variables information efficiently.

The second part applies the Kalman filter, which is trying to estimate the missing observations. With these specifications, Giannone et al. (2008) were able to present a model that included multiple variables that were possible to insert in different periods. In other words, it was possible to include new variables as soon as their data was available.

Giannone et al. (2008) model's results suggested that the more information sets where added; the more precise their model's forecasts were. This outcome recommends using as many information sets as possible and principal component analysis for limiting the model's overfitting. In addition to the dynamic factor model, Giannone et al. (2008) proposed a formal representation for the nowcasting's ragged-edge database problem. It had the following characteristics.

$$Projection\left[y_q \middle| \Omega_v^n\right]$$

Equation 1: Nowcasting GDP growth

Nowcasting projection for given quarters q GDP growth y is dependent on the information set Ω, which is published on a monthly basis v. Because the dynamic factor model typically contains various variables, it also includes multiple series n. (Giannone et al., 2008.)

Giannone et al. (2008) assume that information set Ω_v^n consists of two series $[\Omega_v^{n1}, \Omega_v^{n2}]$. Information series Ω_v^{n1} is available with a one-month lag. The other series Ω_v^{n2} possess higher publishing frequency; therefore, it is available without lag. (Giannone et al., 2008.) Another way of examining these two-information series is to define them as "*hard*" and "*soft*" information sets.

Hard information is typically directly measurable data, for example, industrial production (Götz and Knetsch, 2019). Hard information series are produced by government agencies that are under strict policies and fixed publishing schedule. Moreover, their statistics releases are difficult to accelerate since government statistics agencies collect data from multiple companies and other government agencies, for example, tax administrations. To ensure high-quality data, government agencies employ rigorous quality control methods and revisions (Statistics Finland, 2007). However, these methods are quite time-consuming.

Soft information relates to survey or sentiment data, for example, consumer confidence data. This soft information is projecting consumers' sentiment regarding the economic situation. This type of information is usually less time consuming to publish since it is possible to collect it directly from interviews.

Therefore, soft information provides more timely statistics than hard information. (Bańbura, Giannone, Modugno, & Reichlin, 2013; Götz and Knetsch, 2019.) Because of its timeliness, nowcasting literature has extensively studied soft information's abilities.

Giannone et al. (2008) found that survey data had a significant impact on the GDP in-sample forecasts. Later Bańbura and Rünstler (2011) confirmed that survey data could provide additional information regarding GDP growth when there is no official hard information available. Hence, more timely soft information may produce early signals concerning the future GDP growth.

Consequently, the particular quest has been to find sound and appropriate soft data sources, i.e. alternatives for survey and sentiment data. One possible alternative to these traditional data sources is to use Google Trends data. Google Trends data is available in real-time, and it is quite easy to collect.

In addition, as more individuals are using the internet searches, it covers impressively large population. However, because Google Trends data is the property of Google LLC, they can adjust it, as they want. Furthermore, Google Trends data's range is still relatively limited; hence, extensive analysis concerning the long-term economic conditions is difficult to conduct. Despite these shortcomings, early studies seem to support Google Trends data's role as a noteworthy data source.

Some of these studies suggest it is able to generate as accurate statistics as the survey data (Donadelli, 2015; Della Penna & Huang, 2009). Google Trends has also been found to produce somewhat favourable initial nowcasting results (Götz & Knetsch, 2019; Vosen & Schmidt, 2011). More on these and other related Google Trends studies in the following subchapter, which provides a brief overview of Google Trends literature.

2.2 Studies using Google Trends data

The use of internet search data in economic literature started from Ettredge, Gerdes and Karuga (2005) study, where they used it to predict the unemployment rate in the United States. They argued that by using internet searches, individuals expose information regarding their desires, interests and worries. Results suggested that even limited internet search data had a significantly positive relation to the unemployment rate (Ettredge et al., 2005). At the same time, other fields also started to use internet search data in their research. For example, Cooper et al. (2005) used it in a cancer-related study.

Ginsberg et al. (2009) were the first to use specifically Google search data in scientific research, in which they tried to track influenza illness in the United States. However, economic nowcasting started using Google Trends, when Choi and Varian (Choi & Varian, 2009a; Choi & Varian, 2009b) published their first Google Trends research papers.

Choi and Varian combined these early studies in their 2012 paper, in which they studied Google Trend data's ability to predict the current unemployment claims, consumer confidence, travelling, and car sales (Choi & Varian, 2012).

Choi and Varian (2012) had positive results on Google Trends data's ability to predict unemployment claims. Choi and Varian found that Google Trend data implemented models were able to identify a few turning points in the series (Choi & Varian, 2012, 5–6). Furthermore, time series models have known issues predicting turning points from the data, e.g. Hamilton (2011). Pinpointing these turning points is important because, with sound information regarding the current economic situation, policymakers can use the appropriate policy tools.

However, one can question the robustness of these results. Firstly, Choi & Varian (2012) unemployment model's estimation period was relatively short as it ranged from 2004 to 2011. With survey data, short-term forecasters can use more extended estimation periods. Secondly, the study's benchmark model was a simple AR-1 (Choi & Varian, 2012, 5).

In other words, the benchmark model included only the lag values of unemployment claims. With this model specification, the comparison is not reliable as more variables typically produce additional information. For a more decisive analogy, Choi and Varian (2012) could have used survey data as a benchmark for the Google Trends data.

Nevertheless, these results led to further studies using Google Trends data to predict countries unemployment rate. D'Amuri and Marcucci (2009) analyzed an impressive amount of times series models in their research concerning the United States unemployment rate. Moreover, they created a new Google Index indicator by using the search term "jobs". D'Amuri and Marcucci (2009, 17–19) compared these Google Index models to survey data models. Results indicated that Google Index augmented models were the most accurate in predicting the United States unemployment rate (D'Amuri & Marcucci, 2009, 19–20).

There have also been numerous studies with international Google Trends data. Suhoy (2009) studied Israel's Google Trends data and found that it provides useful information about the current economic situation and especially concerning the current unemployment rate. Askitas and Zimmermann (2009) used German Google searches and discovered strong evidence that searches were able to explain the German unemployment rate.

Tuhkuri (2014) examined whether models with Finnish Google search data models could explain the Finnish unemployment rate. According to Tuhkuri, models that were using Google search data outperformed traditional time series models. He also found that Google search data models were especially helpful in identifying turning points in the unemployment rate. (Tuhkuri, 2014, 20.)

Anttonen (2018) studied Euro areas unemployment rate with advanced Bayesian vector autoregressive (BVAR) model. Antonen (2018) also analyzed BVAR using Google search data similar to Tuhkuri (2014). Google search data did not seem to improve initially efficient BVAR model (Anttonen, 2018, 18–19). Anttonen (2018, 21) argues that this was because the first principle component did not capture enough information regarding Google search data.

Like unemployment claims, Choi and Varian (2012) also had favorable results for consumer confidence. They used Google Trends data to forecast Australian consumer confidence and found that it over-performed the baseline (AR–1) model (Choi & Varian, 2012, 7–8).

Similar to Choi and Varian's (2012) paper, there are also other related studies examining Google Trends ability to nowcast consumer confidence. Della Penna and Huang (2009) constructed a consumer confidence index using Google Trends data. They found a strong correlation between their consumer confidence index and two major survey-based indexes, which were the Conference Board Confidence Index (CCI) and the University of Michigan Consumer Sentiment Index (MCSI) (Della Penna & Huang, 2009).

Vosen and Schmidt (2011) analyzed whether Google Trends data could nowcast private consumption in the United States. Their results suggest that Google search data is more accurate in explaining private consumption than the CCI and MCSI indexes (Vosen & Schmidt, 2011, 12). One possible explanation for this result is that survey-based indicators are not able to capture the actual consumption. In turn, they measure only the expected consumption. However, Vosen & Schmidt (2011, 12) note that their study's estimation period was relatively short, i.e. ranging from 2005 to 2009. Later, Vosen & Schmidt (2012) extended Google Trends consumption research to Germany, where they found similar results.

Likewise, Kholodilin, Podstawski and Siliverstovs (2010) studied Google Trend data's ability to nowcast the United States private consumption. In addition to the MSCI and CCI indexes, Kholodilin, Podstawski and Siliverstovs used financial market variables that included different types of interest rates and the S&P 500 stock market index. Results showed that Google Trend data augmented model is indeed able to forecast private consumption in the United States. At the same time, traditional survey and sentiment data were able to produce similar forecasting results. (Kholodilin et al., 2010, 13–14.)

Numerous other consumer-related studies have used Google Trends data in their nowcasting models. Choi and Varian (2009b) examined Google Trends ability to predict home sales in the United States. Models that included the Google Trends data model had significantly better forecasts than the model without them (Choi & Varian, 2009b, 13). Similar to earlier studies, Choi and Varian's (2009b) estimation period were quite short, and models were rather simplistic.

Regardless, Choi and Varian's (2009b) paper encouraged additional nowcasting studies to use Google Trends for predicting housing markets. Wu and Brynjolfsson (2015) studied the United States housing market in national and state level. They argue that because there is no strategic or bargaining situation involved when searching for information, internet searches could be "honest signal" for consumer's preferences and interests (Wu & Brynjolfsson, 2015, 90; Pentland, 2010). In other words, internet searches could reveal consumers' underlying behavior. Wu & Brynjolfsson (2015) results suggest that Google Trends data is associated with future house prices and sales.

There are also few studies for European housing markets. McLaren and Shanbhogue (2011) compared Google Trend data augmented models to models with official statistics in the United Kingdom's housing market. McLaren and Shanbhogue (2011, 135) also emphasize Google data's real-time limitations as the search terms are not in absolute numeric form.

In this case, searches are a random sample of all searches. This kind of random sampling can cause real-time search results to vary on consecutive days. Moreover, this can be particularly problematic with less popular search terms. (McLaren and Shanbhogue, 2011, 135.)

They report that models with Google Trends variables led to lower prediction errors; hence, they provided useful information about the current housing market. (McLaren & Shanbhogue, 2011, 138). Google Trend data also presented similar results for the Netherlands housing market, where the search term "mortgages" was found to correlate with Dutch housing transactions (Veldhuizen, Vogt & Voogt, 2016).

Choi and Varian (2009b & 2012) were the first to study Google Trend data's ability to predict travelling. According to the results, Google search data improved Hong Kong tourist flow predictions (Choi & Varian, 2012). As before, their benchmark model was rudimentary. Besides, the model was analyzed only for in-sample forecasts (Choi & Varian, 2012, 7). Hence, Choi and Varian's (2012) results reliability is under question.

Still, the travelling theme is relevant for countries, which economies are heavily reliant on tourism. One of these countries is Spain, where Artola and Martínez-Galán (2012) examined Google Trend data's ability to nowcast British tourist visiting Spain. Their study suffers from ambiguity as research results only vaguely reported. Moreover, Artola and Martínez-Galán (2012) stated that Google Trends data could produce helpful information about British tourists. However, these results depended on the chosen time series model (Artola & Martínez-Galán, 2012, 26). Therefore, the extrapolation of these results is somewhat limited.

In summary, these previous studies suggest that Google Trends data can provide somewhat useful information concerning current and near-future consumer behavior. However, the estimation period was relatively short in these early studies.

Despite this, there are also currently a growing number of studies, where researchers use Google Trends data to nowcast financial markets and broad macroeconomic variables. Studies regarding the financial markets are examining whether Google Trend data contain information regarding investors' sentiments. In other words, they are trying to find a relationship between investor's attitudes for a particular stock and Google searches.

Preis, Reith and Stanley (2010) were one of the first to study the connection between financial markets and Google Trends data. Furthermore, they found a strong correlation between the S&P 500 stocks trading volume and Google Trend data (Preis et al., 2010). Bank, Larch and Peter (2011) studied Google Trends data's forecasts for German stocks and liquidity.

According to Bank, Larch and Peter's study, Google search reflect uninformed investors interest in German companies, which they found to correlate with stocks trading volume and liquidity (Bank et al., 2011, 263.)

Perlin et al. (2017) studied Google Trends data's ability to forecast international financial markets, which included markets from Australia, Canada, the UK and the USA. Google Trends data was able to forecast financial markets, and it was exceptionally accurate during the 2009 financial crisis. Perlin et al. recommend that Google Trends database should be included in financial research because it provides helpful early signals of decreased equity prices and increased volatility. (Perlin et al., 2017, 466.)

In addition, researchers have tried to nowcast multiple other macroeconomic variables with Google Trends data. Koop and Onorante (2013) studied Google Trends data with nine different macroeconomic variables that included United States inflation and industrial production. Koop and Onorante (2013, 3) argued that Google searches proxy people's "collective wisdom" and therefore, it could be used to nowcast, for example, inflation. Because Koop and Onorante use multiple macroeconomic variables, they have high dimensional data set (Koop & Onorante, 2013, 5).

To solve this issue, Koop & Onorante apply advanced econometric methods, e.g. TVP regression and model switching (Koop & Onorante, 5–8). They conclude that Google Trends data improves overall nowcasting forecasts compared to the benchmark model, which did not include Google data. (Koop & Onorante, 2013, 9–11.) However, Koop & Onorante (2013) are ambiguous about their models' results, making them difficult to interpret.

Similar to the financial market and investor sentiment, there are also studies where the national sentiment is under examination. These papers focus on analyzing the macroeconomic uncertainty through policy-related uncertainty indexes. In his article, Donadelli (2015) studied Google Trend data's use as policy-related uncertainty indicator or index.

Donadelli (2015) used Google Trends data to form a policy-related uncertainty index for the United States macroeconomic situation. Donadelli stated that growth in Google searches regarding the macroeconomic situation is a signal for the uncertainty of the current economic situation (Donadelli, 2015, 802). According to the results, Google data index can produce similar information as other uncertainty indexes, i.e. VIX-index and news-based indexes (Donadelli, 2015, 805). These results indicate that Google Trends data is a relevant indicator of economic uncertainty.

2.3 Nowcasting GDP growth with Google Trends data

These earlier studies seem to suggest that Google Trends data has many useful features and functions for macroeconomic variables. One of the newest applications for Google Trends data is to use it for forecasting country's gross domestic product (GDP). It is well-known that government agencies publish GDP statistics with a significant time lag. Business cycles can change swiftly and suddenly; therefore, it is in central banks and policymaker's interests to have real-time statistics on the current economic situation. There are currently few studies where Google Trends data have considered providing more timely statistics regarding the country's GDP.

Götz and Knetsch (2019) studied Google Trends data's ability to forecast Germany's GDP. To do this, they used simplistic bridge equation models that models are commonplace in central banks. Götz and Knetsch argued that model's simplicity enables transparent examination about Google Trend data's effects. In bridge models, each GDP component has a separate model. Furthermore, Götz and Knetsch assume that these GDP components represent different industry sectors. (Götz and Knetsch, 2019, 46–48.)

Götz and Knetsch's industry models include short-term indicators, i.e. timely information concerning the particular industry. Therefore, short-term information is being "bridged" to the GDP estimation. Götz and Knetsch divide these short-term indicators into soft and hard indicators. Former relates to the survey data and the latter, for example, data on the industrial production. Given its properties, they consider Google Trends data as a soft indicator. (Götz & Knetsch, 2019, 46–48.)

Moreover, Vosen and Schmidt (2011) previously stated that Google Trends data provides more accurate predictions about the current consumption than the survey data. Hence, this further suggests that Google data could be a possible alternative for traditional survey data.

Götz and Knetsch estimated bridge models using European Central Banks (ECB) search data. This ECB data differs from the publicly available Google Trends data. Publicly available data includes more categories than ECB data. However, ECB data is normalized to begin from one when the public data starts from zero. Götz & Knetsch, 2019, 49.) Götz and Knetsch also argue that the ECB data is more accurate as "the random samples on which the data are based are much smaller" (Götz & Knetsch, 2019, 49).

Compared to other traditional data sources, Google data is typically highly dimensional, i.e. there are multiple variables for a limited amount of time-series data. This particular property calls for meticulous variable selection. For identifying the most efficient Google variables, Götz and Knetsch used multiple variable selection methods (Götz and Knetsch, 2019, 50–51).

These included partial least squares (PLS), shrinkage, principal component analysis (PCA), boosting, selection operator (LASSO) and a few "ad hoc" approaches. These "ad hoc" methods were the most simplistic as one of them included just using "common sense". (Götz and Knetsch, 2019, 50–51.)

In other words, they were selecting search terms that they thought to have actual economic relation with GDP. The "ad hoc" method also utilized Google correlate service, which singles out variables that are moving in the same direction. (Götz and Knetsch, 2019, 50–51.)

After the reduction of dimensionality, Götz and Knetsch conducted nowcasting forecasts for three different model specifications. Götz and Knetsch compared these models in two parts. First, they compared models, which included hard indicators, Google Trends data and survey data to the benchmark model. Benchmark model included only the hard indicators and the traditional survey data. Second, they compared the benchmark models forecast to models, which excluded the survey data. (Götz & Knetsch, 2019, 51–55.)

Their estimation period spanned years 1991–2016. Google Trends data is available since 2004; ergo Google variables spanned years 2004–2016. (Götz & Knetsch, 2019, 51.) Known issue when studying GDP growth is the ragged-edge database problem. Götz and Knetsch (2019, 52) solved this issue by estimating every dataset that was not available in the forecasting period. Regardless, Götz and Knetsch (2019, 53–55) analyzed nowcasting models by their root mean squared forecast error (RMSFE) results, which is a standard method to examine time-series forecasts. In other words, they compared Google augmented models RMFSFE results to the benchmark models results.

Results suggest that Google Trends data is capable of providing some additional information regarding the German manufacturing, hotel and mining sector. However, models that included both the Google and survey data suffered forecast accuracy losses in construction and net tax sectors. (Götz & Knetsch, 2019, 53–54.) According to Götz and Knetsch, models that included only Google Trends data as a soft indicator produced low RMSFE results in the long and mid-term. Nevertheless, the benchmark model exceeded Google augmented models in the near-term. It seems that Google Trends data is missing some valuable information about the near-term. (Götz & Knetsch, 2019, 53–54.)

In summary, Google Trends data can provide additional information when there is no official survey data available. However, official survey data is available monthly, which implies that Google Trends data's gains are somewhat limited. Similar to Götz and Knetsch paper, there are other central bank-related studies concerning Google Trend data's use in forecasting GDP.

The most recent working paper by Ferrara and Simoni (2019) examines Google Trends data's effectiveness to nowcast euro areas GDP growth. For achieving this, they used bridge equation models to study GDP growth in Germany, France, Italy, Netherlands, Belgium, and Spain. (Ferrara & Simoni, 2019, 1–3.)

Ferrara and Simoni used both the hard and soft information sets in their bridge models. The hard information was the euro area's industrial production. Soft information was the euro areas sentiment index, which is a survey index from various industry sectors. Ferrara's and Simoni's also used Google Trends data set, which had a significant number of variables, 1776 in total. (Ferrara &

Simoni, 2019, 1–3.) This number of variables naturally leads to high dimensionality. However, dimensionality is possible to reduce with different variable selection methods.

To do this, Ferrara and Simoni employed a machine-learning technique called Ridge regression and Sure Independence Screening method (SIS). SIS method's objective is to find variables that provide the most significant correlations with the GDP growth. Ferrara and Simoni solved nowcasting's ragged edge database problem by constructing 13 different models for each week of the quarter. (Ferrara & Simoni, 2019, 6–7, 10.)

Ferrara and Simoni compared these models based on their RMSFE results. According to Ferrara & Simoni, models that included Google Trends data were able to produce valuable information for nowcasting GDP. However, this information was valid for only the first four weeks of the quarter. For the fifth week, official survey statistics were able to outperform the Google Trend model.

They conclude that Google Trend data forecasts are the most useful when there is no official data available, i.e. survey data. (Ferrara & Simoni, 2019, 14–16, 21.) Still, it is possible to question further Ferrara & Simoni's results, as models without Google variables were able to get the lowest RMSFE results. In other words, the most accurate forecasting models did not include Google Trend variables. (Ferrara & Simoni, 2019, 15.) This result undermines Google Trend status as an alternative for traditional survey data.

In conclusion, earlier GDP studies using Google Trends have had quite mixed results. Studies suggest Google data could provide some additional information concerning the GDP growth. In Germany, this information seems to relate to its manufacturing sector (Götz & Knetsch, 2019, 53–54). German's relatively large manufacturing sector could explain this relation, i.e. the automotive industry.

Furthermore, Google data's additional information was found to be particularly potent in the first four weeks (Ferrara & Simoni, 2019, 14–16). One can challenge the reliability of these results because the model that contained the survey data had the most accurate forecasts. Illuminated by these early results, this master's thesis attempts to examine whether Google Trends data could nowcast Germany or Finland's GDP growth. Following sections describe this examination in detail.

Table 1: Studies with Google Trends

Studies with Google Trends	Country	Economic variable(s)	Key result(s)
Choi & Varian (2009a, 2009b and 2012)	USA, Hong Kong and Australia	Multiple different economic variables	Google Trends provided useful information about unemployment claims, consumer confidence, home sales and travelling.
D'amuri & Marcucci (2009)	United States	Unemployment rate	Google model had the most accurate forecasts
Suhoy (2009)	Israel	Unemployment rate	Google searches provided additional information about unemployment
Askitas & Zimmermann (2009)	Germany	Unemployment rate	Google searches were able to explain unemployment rate
Tuhkuri (2014)	Finland	Unemployment rate	Google Trends model generated the most accurate forecasts
Anttonen (2018)	Euro area	Unemployment rate	Google search data did not improve initially efficient BVAR model
Huang & Della Penna (2009)	United States	Consumer confidence	Google searches had substantial correlation with consumer confidence
Vosen & Schmidt (2011 and 2012)	USA and Germany	Consumption	Google Trends data is capable to explain private consumption
Kholodilin, Podstawski & Siliverstovs (2010)	United States	Consumption	Google models produced similar results as consumer confidence models
Wu & Brynjolfsson (2015)	United States	Housing market	Google searches were found to be linked with future house sales and prices
McLaren & Shanbhogue (2011)	United Kingdom	Housing market	Google Trends models had the most accurate forecasts
Veldhuizen, Vogt & Voogt (2016)	Netherlands	Housing market	The search term "mortgages" had a significant correlation with housing transactions
Artola & Martínez-Galán (2012)	Spain	Travelling	Google Trends data was able to generate information about future tourists
Preis, Reith & Stanley (2010)	United States	Stocks trading volume	Found a significant correlation between S&P 500 stocks trading volume and Google Trends data
Bank, Larch & Peter (2011)	Germany	Stocks and liquidity	Google Trends data was found to correlate with stocks trading volume and liquidity
Perlin, Caldeira, Santos & Pontuschka (2017)	Australia, Canada, UK and USA	Financial markets	Google Trends produced additional information about financial markets
Koop & Onorante (2013)	United States	Nine different macroeconomic variables	Implementing Google Trends data improves forecasting accuracy
Donadelli (2015)	United States	Policy-related uncertainty	Google data generated similar information as other uncertainty indexes
Götz and Knetsch (2019)	Germany	GDP	Survey data outperformed Google Trends data.
Ferrara & Simoni (2019)	Euro area	GDP	The most accurate models did not include Google Trends data

3 DATA

This master's thesis uses three different types of data to study the intriguing topic of nowcasting GDP growth using Google Trends. These data sources include both Finland's and Germany's official GDP statistics, consumer survey statistics and Google Trends data. For this purpose, the data section has a three-part structure. The first part consists of a discussion concerning the countries official GDP data series. The second part focuses on describing countries consumer survey data. The third and the final section examines the properties of the Google Trends data.

3.1 Gross Domestic Product (GDP) data

This master's thesis utilizes GDP volume data measured in changes compared to the previous quarter to measure a country's GDP growth. The GDP data was also seasonally, and working day adjusted. Statistics Finland publishes Finland's GDP data as part of their national accounts' statistics. OECD publishes a wide range of economic data, one of which is Germany's GDP data. These published GDP statistics depicts how countries GDP has fluctuated through time, i.e. the GDP growth. Figures 1 & 2 illustrate Finland and Germany's quarterly GDP growth from 2004 to 2018.

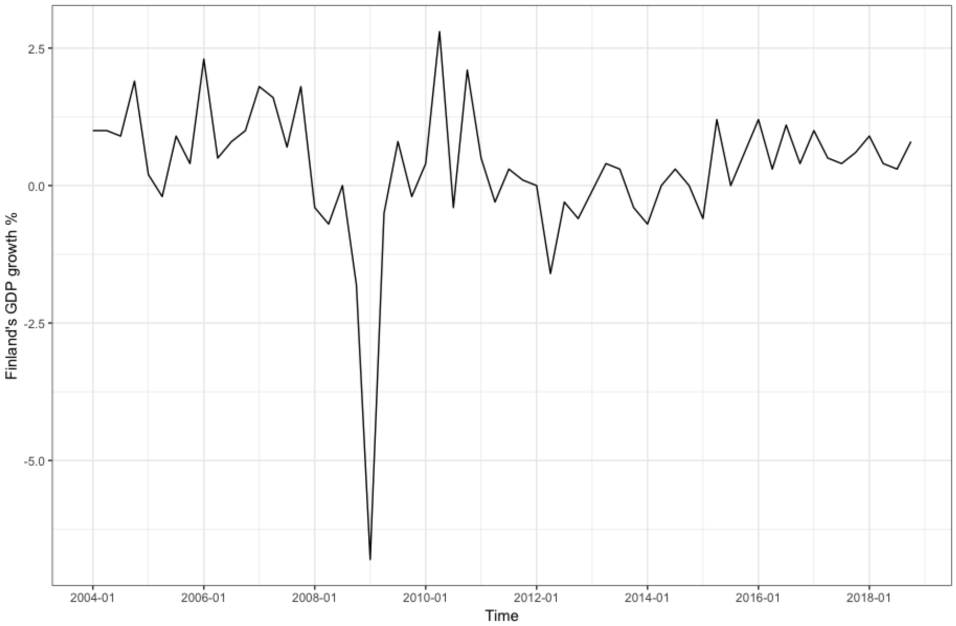


Figure 1: Finland's quarterly GDP growth from 2004 to 2018 (Statistics Finland, 2019b)

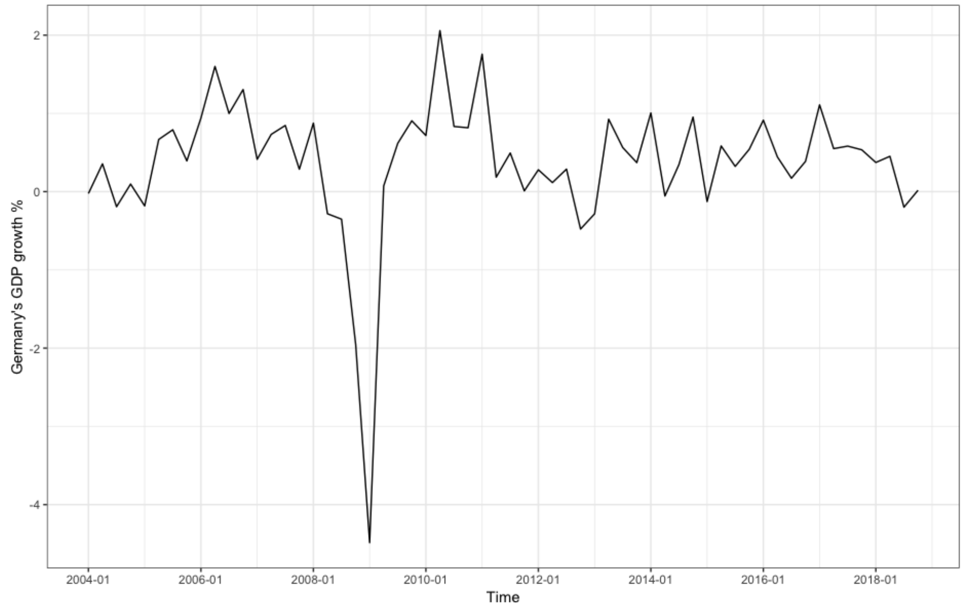


Figure 2: Germany's quarterly GDP growth from 2004 to 2018 (OECD, 2019)

Compared to the monthly data series, both of the official GDP statistics had a reporting lag of two months. For example, in 2019, Statistics Finland's released fourth-quarter GDP statistics at the end of February. On the other hand, the official consumer survey data and Google Trends data are available monthly; hence, they do not have a similar time lag.

3.2 Consumer Confidence data

Consumer confidence survey has been one of the most used soft data sources in nowcasting models, e.g. Bańbura, Giannone & Reichlin (2010). Consequently, this master's thesis uses consumer confidence survey as a benchmark for a typical soft data. In other words, this thesis regards Google Trends data as an alternative to Finland and Germany's consumer survey data.

As previously stated, Statistics Finland publishes Finland's consumer survey monthly. Statistics Finland constructs these surveys by interviewing over 2000 individuals living in Finland. Interviews measures individuals' confidence and expectations about Finland's or their own economy (Statistics Finland, 2017.)

More specifically, this study applies Finland's consumer confidence data regarding consumer's confidence in their own economy. This data series is also a logical choice because its properties are similar to Google Trends data. In short, they are both short-term proxies for consumers' behaviour. Figure 3 depicts consumers' confidence data in their own economy from 2004 to 2018.

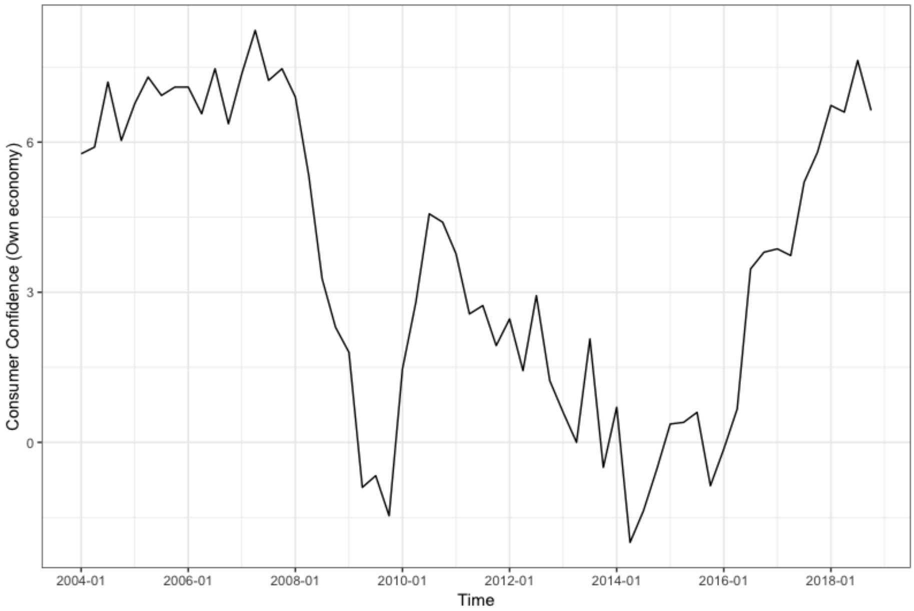


Figure 3: Finland’s monthly consumer confidence in their own economy from 2004 to 2018 (Statistics Finland, 2019a)

Figure 3 implies that consumer confidence decreased in during the 2008 financial crises. Confidence briefly recovered shortly after the crises. This recovery was short-lived, and in 2014, Finland’s consumers had the lowest confidence about their economy. However, after this low point, the confidence rose steadily and finally reaching its pre-crisis levels in 2018.

For Germany, this master’s thesis uses European Commissions consumer survey data. More explicitly, consumer confidence data concerning consumers the financial situation over the last 12 months, which is similar to earlier Finnish confidence data.

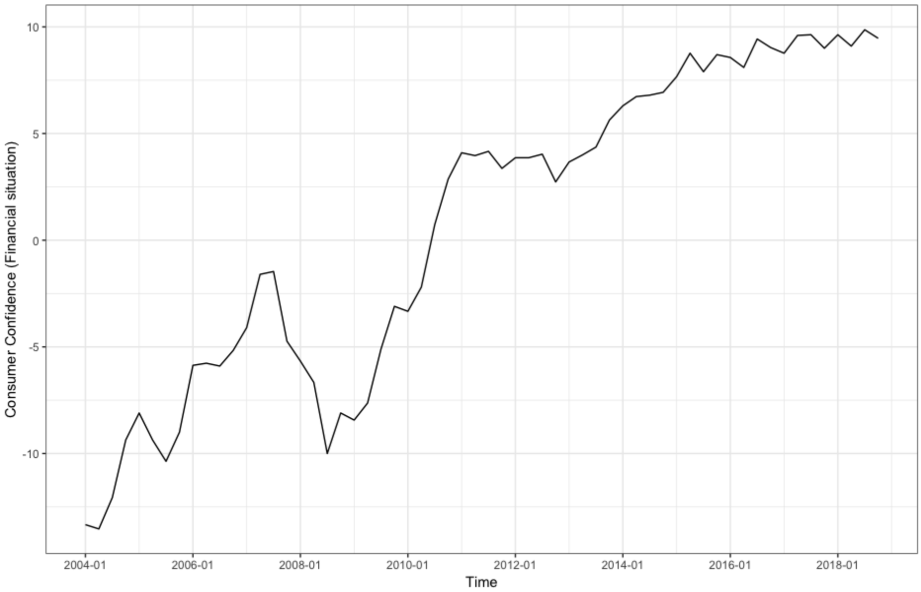


Figure 4: Germany’s monthly consumer confidence from 2004 to 2018 (European Commission, 2019)

Unlike in Finland, the confidence of German consumers has mostly strengthened in the last fourteen years. According to figure 3, the only significant drop in confidence occurred during the global financial crises. Overall, German consumers seem to be highly confident regarding their financial situation.

3.3 Google Trends data

Google Trends data are available on Google's website, which allows users to type in different search terms. Moreover, users can specify search terms for different geographical levels. For example, the website reports Finnish search term results for both the country and municipality levels.

Google Trend data website also enables users to specify the range of the search terms; for example, users can set search data to begin from the past hour. The maximum range for the Google Trend data spans from the year 2004 to the present day. However, this maximum range is only available in the form of monthly data. In addition, the website provides related topics and queries; in the case of GDP, these consists of other macroeconomic factors such as inflation and human development index.

It is worth noticing that the website does not publish the search data in absolute numerical form; instead, it is available in the form of a search ratio. Following equation 2 provides a formal depiction of the search ratio.

$$SR_{ig} = \left(\frac{s_{ig}}{\sum_{i=1}^{\mathrm{N}} \sum_{g=1}^{G} s_{ig}} \right) * 100$$

$$i = 1, \ldots, \mathrm{N}$$
$$g = 1, \ldots, \mathrm{G}$$

Equation 2: Search ratio

Search ratio SR for a search term i in a geographical area is possible to present as a division. In it, search term i in a geographical area g is divided by sum of all the search terms n in a particular geographical area. Finally, the result of this division is then multiplied by 100. (Choi & Varian, 2012; Google, 2019b.)

In other words, the search ratio ranges from 100 to zero, where 100 states that search term is relatively popular in the chosen region. According to Google, this normalization allows for a smoother comparison between search terms as search volumes vary between different countries. (Google, 2019b.)

Google divides Trends data into non-real-time data and real-time data. Non-real-time data covers more ground as it is a random sample of Google search, which is possible to collect since the year 2004. Real-time data is more frequently, and the random sample is possible to collect from the past week. (Google, 2019b.)

Google Trends website provides data only for popular search terms. Furthermore, Google states that data do not include duplicate "searches from the same person over a short period of time". This duplicate search term control reduces the possibility of people deliberately affecting search terms popularity. Google also specifies that Trends data include only search terms without special characters or apostrophes. (Google, 2019b.)

Since August 2008, Google has classified different search terms into various categories (Google, 2008). In other words, if the user searches "apple", it could mean the fruit or the computer company. Google assigns these search terms into a specific category by using probabilities; for example, search term "apple" into the Food & Drink category (Google, 2019c; Choi and Varian, 2012, 4).

Google Trends uses 27 broad categories that include categories covering, for example, searches about News, Shopping and Jobs. Also, Google further divides these broad search term categories into over 1400 subcategories, which vary from specific scripting languages to Gothic subcultures.

One advantage of these categories is that the researcher does not need to worry about language-specific search terms. Still, with an abundance of possible categories and variables, the researcher needs to proceed with caution to determine which subcategories are relevant for GDP growth. This master's thesis follows Götz and Knetsch (2019) paper to select appropriate initial subcategories. However, the sensitive subject's category is excluded from this thesis because it was not available on the Google Trends website. These initial subcategories are in appendix 1.

As presented in the appendix, there are over 180 initial Google Trends subcategories (i.e. variables) from 16 different broad categories. Because of this, the data series is highly dimensional. Consequently, this study applies modern dimension reduction methods, which are similar to Götz and Knetsch (2019) paper. With these dimension methods, initial 180 subcategories were compressed into 16 different broad categories. Table 2 shows these broad categories.

Table 2: Compressed Google Trends broad categories

Autos & Vehicles	Beauty & Fitness	Business & Industrial	Computers & Electronics
Food & Drink	Health	Home & Garden	Internet
Investing	Jobs	Law	News
Real Estate	Shopping	Sports	Travel

The initial categories were in monthly form, and they ranged from January 2004 to March 2019. Moreover, this master's thesis possesses the longest possible range of Google Trends data available in early 2019.

Figures 4 & 5 present both Finland and Germany's Food & Drink category data. The figures also illustrate the Food & Drink category against countries' GDP and consumer confidence data. In these figures, the monthly data series were aggregated to quarterly levels by calculating their three-month averages. In addition, Food & Drink category variables were compressed into a single common factor by the principal component analysis (PCA).

Furthermore, in this analysis, a common factor was created by selecting the first principal component. As suggested by Giannone et al. (2008, 668), common factors are a good approximation for high dimensional data sets. The following section 4 discusses the principal component method in greater detail. Nevertheless, subsequent figures 4 and 5 show the Food & Drink category's first principal component (PC1) and countries' GDP growth.

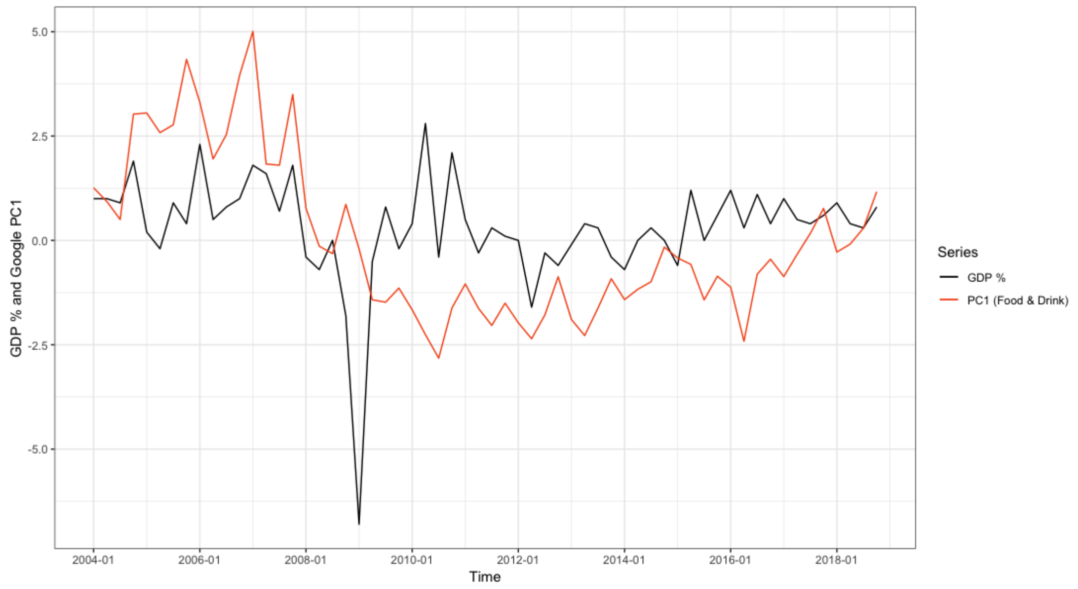


Figure 5: Finland's GDP growth and first principal component for the Food & Drink category

Figure 5 implies that the number of Food & Drink category related search terms have been varying quite substantially. The most noticeable trend is a large number of Food & Drink related searches in the pre-financial crises. It is also interesting that these searches decreased in amidst of 2008 financial crisis. Furthermore, it was a long-term decrease in Food & Drink related searches.

Searches for Food & Drink related search terms might have initially increased with individual's better internet access and people's interests eating at restaurants, as the Food & Drink category includes search terms for restaurants. For a more specific description of the Food & Drink category is in appendix 1. However, in 2008, news about the financial crisis greatly affected people's incentives for saving and eating at home.

In addition to the initial short-term effect, the financial crisis had a long-term effect, and people's interest in Food & Drink related continued to stay relatively low. Searches were able to reach their pre-crisis levels as late as 2018. These developments in Food & Drink searches could be a reflection of Finland economy's structural change, which began from the 2008 financial crises.

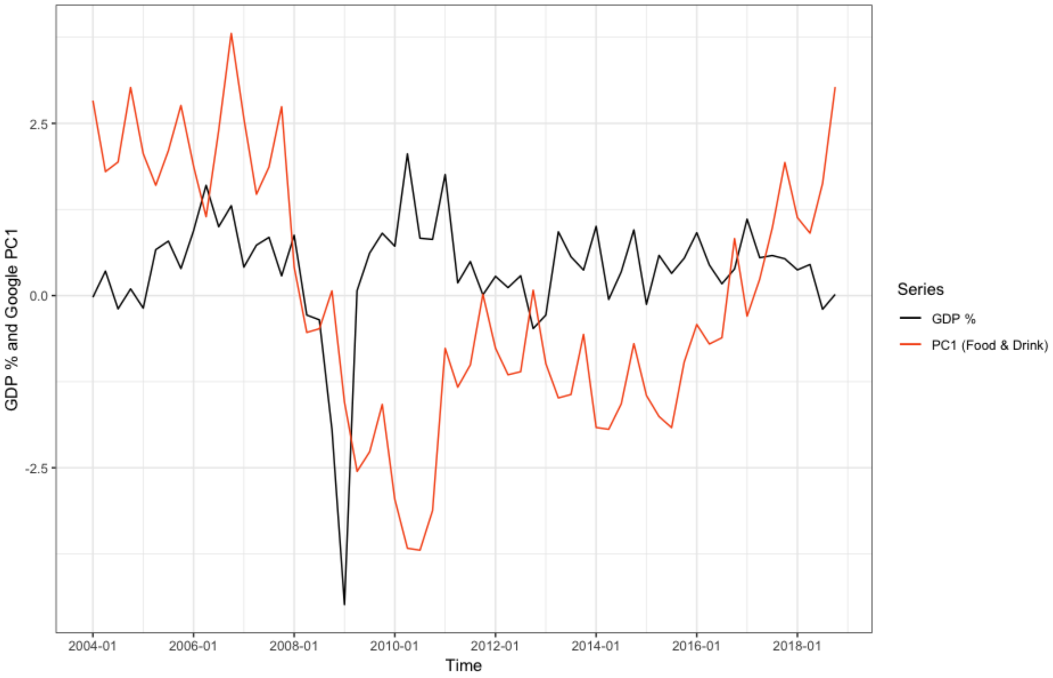


Figure 6: Germany's GDP growth and first principal component for the Food & Drink category

Similar to Finland's results, Germany's Food & Drink category searches have a relatively high variance. People were doing much Food & Drink related searches before the financial crises. These searches decreased in the aftermath of the crises. It could be that financial crises changed people's incentives to eat more at home. Germany's Food & Drink searches have increased and decreased more rapidly and searches were able to catch up with GDP a lot earlier compared to Finland.

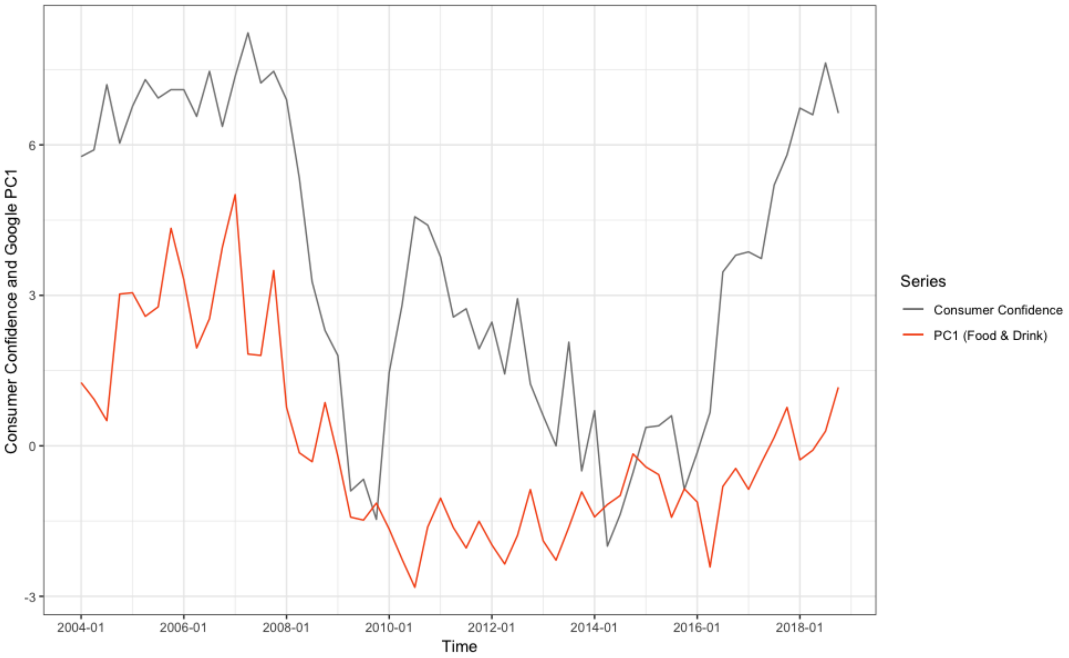


Figure 7: Finland's consumer confidence and first principal component for the Food & Drink category

As seen in figure 4, Google search terms for food & drinks and Finland's consumer confidence seem to be opposite images of each other. When consumer confidence is relatively low, searches for food & drink are high. In other words, people search for food & drink when they are not confident about their economy. Moreover, the Food & Drink category also includes search terms for alcoholic beverages. It could be that when the confidence to own economy is low people are seeking relief from alcohol.

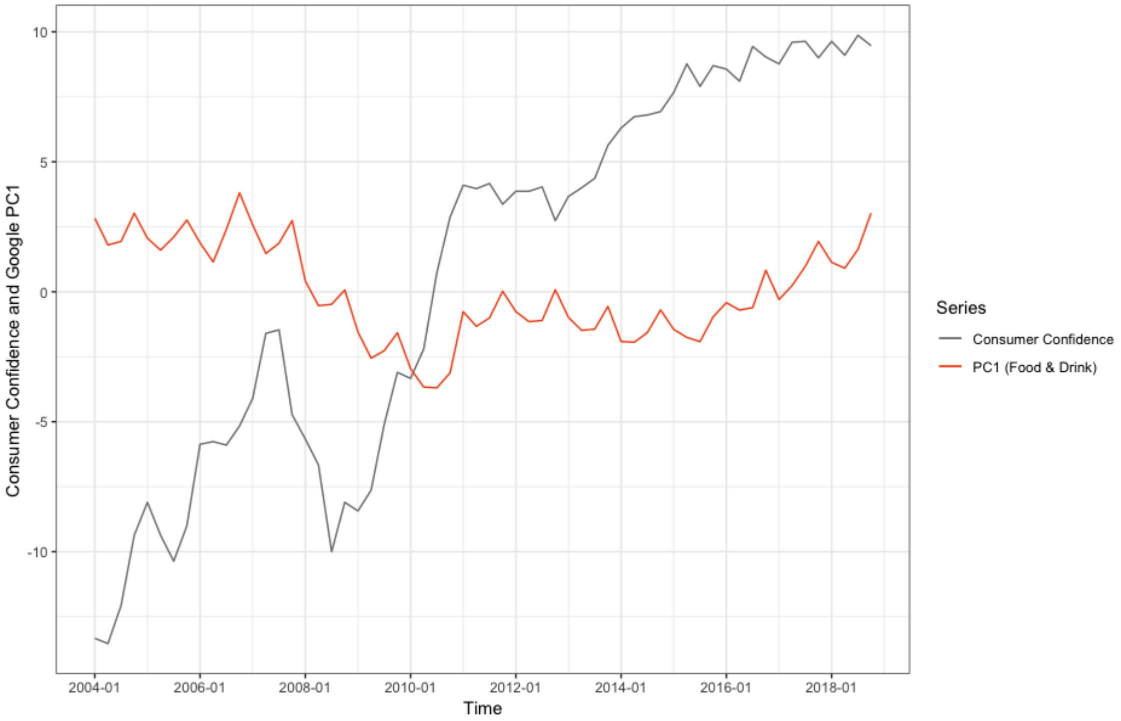
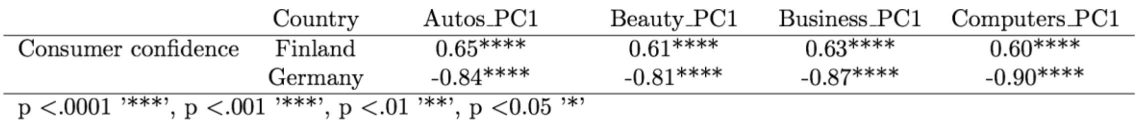


Figure 8: Germany's consumer confidence and first principal component for the Food & Drink category

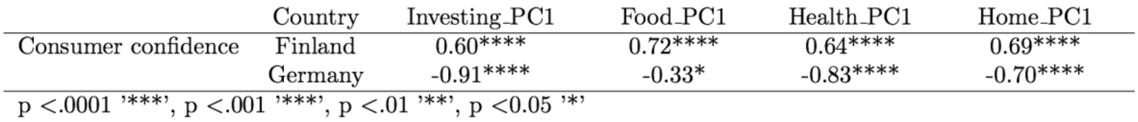
Figure 8 shows that in Germany, there is not as clear a link between consumers and Food & Drink searches as in Finland. This divergence is because German consumers have experienced constant improvements in their financial situations, which have led to higher consumer confidence. The next table 3 describes how other Google category PC1 components relate to countries consumer confidence. Namely, table 3 depicts Google categories correlations against Finland and Germany's consumer confidence.

Table 3: Google categories correlations with consumer confidence[2]



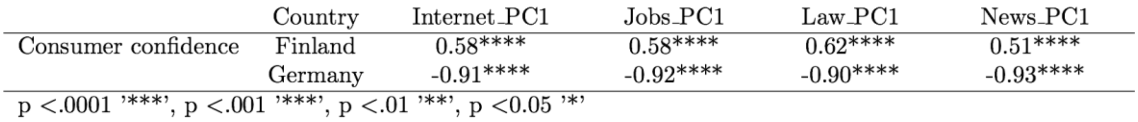
	Country	Autos_PC1	Beauty_PC1	Business_PC1	Computers_PC1
Consumer confidence	Finland	0.65****	0.61****	0.63****	0.60****
	Germany	-0.84****	-0.81****	-0.87****	-0.90****

p <.0001 ’***’, p <.001 ’***’, p <.01 ’**’, p <0.05 ’*’



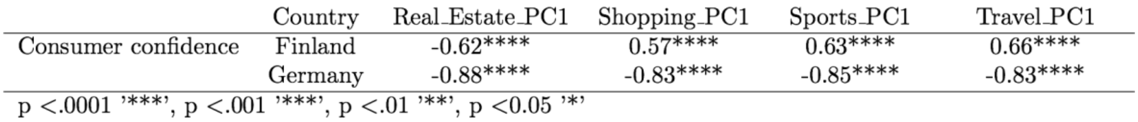
	Country	Investing_PC1	Food_PC1	Health_PC1	Home_PC1
Consumer confidence	Finland	0.60****	0.72****	0.64****	0.69****
	Germany	-0.91****	-0.33*	-0.83****	-0.70****

p <.0001 ’***’, p <.001 ’***’, p <.01 ’**’, p <0.05 ’*’



	Country	Internet_PC1	Jobs_PC1	Law_PC1	News_PC1
Consumer confidence	Finland	0.58****	0.58****	0.62****	0.51****
	Germany	-0.91****	-0.92****	-0.90****	-0.93****

p <.0001 ’***’, p <.001 ’***’, p <.01 ’**’, p <0.05 ’*’



	Country	Real_Estate_PC1	Shopping_PC1	Sports_PC1	Travel_PC1
Consumer confidence	Finland	-0.62****	0.57****	0.63****	0.66****
	Germany	-0.88****	-0.83****	-0.85****	-0.83****

p <.0001 ’***’, p <.001 ’***’, p <.01 ’**’, p <0.05 ’*’

According to table 3, most of the Google categories correlate positively with Finland’s consumer confidence survey. It could be that when consumer confidence is relatively high, Finland’s people are searching for more information. This information could be about nutrition, shopping, or travelling.

However, all of Germany’s Google categories were highly negatively correlated with German consumer confidence. In other words, when German people were most confident about their financial situation, they were using less of their time searching for information. Furthermore, these significant correlations provide some confirmation to Huang & Della Penna (2009) earlier paper regarding Google Trends correlation with consumer confidence.

In summary, table 3 results suggest that in both countries, Google Trends categories share a significant relationship with the consumer confidence data. In Finland, this relationship is mostly positive and in Germany is negative. The subsequent section uses these data sources to nowcast both Finland’s and Germany’s GDP growth.

[2] It is worth noting that principal components have a property, which can lead to a “wrong” initial sign. Principal components were tested against summed Google categories to verify the correct correlation sign. These tests releveled the fact that initial signs were incorrect as principal components correlated negatively with summed Google categories. Thus, this master’s thesis had to correct these correlations. Table 3 displays these adjusted correlations.

4 METHODS

Before this master's thesis can start to discuss or conduct any prominent nowcasting analysis, Google Trend data's high dimensionality properties demand further assessment. For reducing high dimensionality, Götz and Knetsch (2019) used seven different methods that included dimension reduction, shrinkage and a few ad hoc approaches. Ferrara and Simoni (2019) applied both Sure Independence and Ridge methods. This study uses similar methods to mitigate Google Trends data's high dimensionality property, i.e. dimension reduction methods and variable selection method.

4.1 Dimension reduction methods

This master's thesis applied two different dimension reduction methods. The general idea in dimension reduction methods is to reduce the number of predictors by transforming them, for example, into common factors. These common factors or linear combinations can be formally defined with original predictors $X_{1,}\dots,X_{p}$ and constants $\phi_{1m},\dots\phi_{pm}$. (James, Tibshirani, Witten, & Hastie, 2013, 229.)

(3) $$Common\ factor_m = \sum_{j=1}^{p} \phi_{jm} X_j \qquad m = 1,\dots,\mathrm{M}$$

Dimension reduction methods objective is to find the optimal values for constants ϕ_{jm}. These methods reduce the number of predictors to a $Common\ factor_m$. These common factors can then be implemented into linear regression, which can be estimated using the ordinary least squares (OLS). (James, Tibshirani, Witten, & Hastie, 2013, 229.)

(4) $$y_i = \theta_0 + \sum_{m=1}^{M} \theta_m Common\ factor_{im} + \epsilon_t \qquad i = 1,\dots,\mathrm{N} \quad m = 1,\dots,\mathrm{M}$$

Now, due to dimension reduction methods regression has a fewer prediction as $M < p$. Furthermore, instead of $p + 1$ predictors regression has only $M + 1$ predictors. (James, Tibshirani, Witten, & Hastie, 2013, 229.)

One method to find these optimal constants ϕ_{pm} is to use *Principal component analysis* (PCA). PCA's way to reduce dimensionality is by maximizing variance. (Götz & Knetsch, 2019, 56). Equation 5 illustrates this variance maximization as an optimization problem.

(5) $$maximize\{\phi_m^{\mathrm{T}} \Sigma \phi_m\} \qquad constraint\ \phi_m^{\mathrm{T}} \phi_m = 1$$

Equation 5 states that variance is maximized with constants transposed vector ϕ_m^{T}, the covariance matrix Σ of predictors X_j and constant vector ϕ_m. Also, optimization has a normalization constraint. This constraint enables the real maximum solution. (Jolliffe, 2002, 2–5.)

Moreover, this optimization is possible to solve with Lagrange multiplier and eigen decomposition (Jolliffe, 2002, 5 & James, Tibshirani, Witten & Hastie, 2013, 376). Following equations 6 and 7 illustrate Lagrange multiplier optimization.

$$\phi_m^{\mathrm{T}}\Sigma\phi_m - \lambda(\phi_m^{\mathrm{T}}\phi_m - 1) \tag{6}$$

As seen in equation 6, the optimization problem is constrained with Lagrange multiplier λ.

$$(\Sigma - \lambda I_p)\phi_m = 0 \tag{7}$$

Equation 7 is the result of differencing equation 6 with respect to ϕ_m. What is more, it now includes an identity matrix I_p. Now, λ represents the eigenvalue of the covariance matrix Σ. Furthermore, ϕ_m is the eigenvector. (Jolliffe, 2002, 5.) These eigenvectors, also called loadings, explains a different amount of the variance (James, Tibshirani, Witten & Hastie, 2013, 231).

As described in equation 3, the common factors are generated from the constants or component loadings that multiplied with the original predictors' X_j values. In the principal component analysis, the first principal component explains the most substantial amount of the series variance. The second principal component explains the second-largest amount of the series variance orthogonal of the first component. (James, Tibshirani, Witten & Hastie, 2013, 231–232.)

Because the first principal component is defined to explain the most substantial amount of the series variance, it can be used to summarize data (James, Tibshirani, Witten & Hastie, 2013, 379). Therefore, this master's thesis uses the first principal component as a common factor, which used to reduce Google data's high dimensionality. There are also other dimensionality reducing methods.

The other prevalent dimension reduction method is *Partial least squares* (PLS), which is similarly transforming original predictors into new common factors. In the PLS method, common factors are constructed by considering the relationship between the original prediction variables $X_1,\ldots,X_p$ and the dependent variable y_i. (Götz & Knetsch, 2019, 56; James, Tibshirani, Witten & Hastie, 2013, 237.) More formally, PLS decomposes prediction variables and dependent variable into two parts, which are found in the following equations 8 and 9.

$$X_p = T\phi_m^{\mathrm{T}} + E \tag{8}$$

$$y_i = UQ^T + F \tag{9}$$

In equations 8 and 9 T and U are the PLS components. Furthermore, E and F terms are the residuals. Similar to PCA, ϕ_m^{T} are the predictor's X_p loadings. However, in PLS, the dependent variable y_i has also loadings Q^T. PLS method is trying to maximize the covariance between T and U. This maximization is conducted through their weights w and c. (Rosipal & Krämer, 2005, 35–36.)

(10) $$[cov(t,u)]^2 = maximize[cov(Xw,Yc)]^2$$

The maximization problem described in equation 10 is solved with different algorithms that follow similar iterative processes. These algorithms produce a similar result, which is shown in equation 11. (Rospial & Krämer, 2005, 35–36.) This master's thesis used the Kernel algorithm.

(11) $$X^TYY^TXw = \lambda w$$

Algorithms form the first component for the prediction variable X_p, which is in the following equation 12. (Rospial & Krämer, 2005, 35–36.) In PLS, the first component can also be formed for the dependent variable y_i.

(12) $$t = Xw$$

Similar to equation 3, the first component is formed by multiplying the predictor variable X with the eigenvector w (Rospial & Krämer, 2005, 35–36). In other words, the PLS concept to construct the first common factor is somewhat similar to PCA. Likewise, after constructing these factors, they can also be implemented into linear regression.

However, to find optimal factors, PLS gives more weight on the original X_j predictors that are highly correlated with the dependent variable y_i. In addition, when estimating the second common factor, the variables are adjusted by the first common factor. In other words, the second common factor uses only the residuals of the first common factor. This process continues to follow through all of the factors. (Götz & Knetsch, 2019, 56; James, Tibshirani, Witten & Hastie, 2013, 237.)

PLS method's benefit is that it is better in explaining the relationship between the prediction variables and dependent variable. However, according to James, Tibshirani, Witten & Hastie (2013, 237), PLS results are also exposed to greater variance. Nevertheless, PCA and PLS methods advantage is also that they are optimal methods even if the data series are highly correlated (James, Tibshirani, Witten, & Hastie, 2013; Stock & Watson, 2002). Because Google Trend categories are related to one another, they are expected to be highly correlated this, in turn, would produce spurious regressions. Therefore, dimension reduction methods are proper tools to mitigate also this issue.

4.2 Shrinkage method

Shrinkage is a method that includes all the p predictors from the data. Similar to ordinary least squares (OLS), shrinkage methods are trying to create a tight fit of the data by reducing the sum of squared residuals. However, shrinkage methods are subject to a shrinkage penalty or a constraint, which shrinks regression coefficients towards zero. (James, Tibshirani, Witten & Hastie, 2013, 214–215, 221.) In this manner, shrinkage can allow for more efficient use of the data. One of the most used shrinkage methods is LASSO.

LASSO or *least absolute shrinkage and selection operator* is a method, which is shrinking the initial regression predictors by punishing its coefficients. Next equation displays how LASSO adjusts its regression coefficients. (James, Tibshirani, Witten & Hastie, 2013, 219.)

$$(13) \qquad \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_1 x_{ij})^2) + \lambda \sum_{j=1}^{p}\left|\beta_j\right|$$

As noticed in equation 13, LASSO uses a tuning parameter λ to punish its regression coefficients. When the tuning parameter increases, the coefficient decrease. In other words, if the tuning parameter is equal to zero, the model includes all the coefficients at which point it is a typical OLS regression. (James, Tibshirani, Witten & Hastie, 2013, 219.)

Furthermore, LASSO is trying to identify coefficients that produce the lowest regression residual. This identification is possible to present as a minimization problem, which is in the following equation 14. (James, Tibshirani, Witten & Hastie, 2013, 220.)

$$(14) \quad minimize \left\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_1 x_{ij})^2\right\} \quad subject\ to\ constraint\ \sum_{j=1}^{p}\left|\beta_j\right| \leq s$$

Equation 14 states that LASSO is minimizing the regression residual. However, this minimization is under a constraint. If s is relatively small, minimization produces only a few coefficients (James, Tibshirani, Witten & Hastie, 2013, 221). To find the optimal s or tuning parameter λ, one can use the cross-validation method in which the parameter with the lowest cross-validation error is chosen (James, Tibshirani, Witten & Hastie, 2013, 176, 227).

Thus, the LASSO method selects the coefficients that minimize regression residual, and meanwhile, the number of variables in the regression decreases. In other words, this method allows the researcher to use high dimensional data set in a relatively efficient manner.

4.3 Nowcasting exercise and models

This master's thesis constructed *pseudo-out-of-sample* forecasting exercise to examine Google Trends data's nowcasting abilities. The purpose of this exercise was to simulate a real-world nowcasting situation (Stock & Watson, 2008, 1). This sort of pseudo-out-of-sample simulation is possible to estimate with two different strategies. The first strategy is *"fixed" rolling window* in which the sample size does not change. The second strategy is a *recursive* or *expanding window,* where the initial sample is expanding (Stock & Watson, 2008, 3–4). The first strategy demands a relatively long data series. However, because Google Trends is available since 2004, this study uses expanding window nowcasting strategy.

In this master's thesis, the initial sample size was 20 % of the entire data, i.e. 12 periods. This initial sample size a rather short, which is because of the length of the Google Trends data, i.e. 60 periods. The forecasting exercise used both the most recent monthly data and aggregated monthly data to match the monthly data to the quarterly data. With these specifications, this thesis used following forecasting models that were estimated using ordinary least squares (OLS). Gradual description of the pseudo-out-of-sample exercise is in Appendix 2.

$$GDP_t = \beta_0 + \beta_1 GDP_{t-1} + \varepsilon_t \qquad t = 1, \dots, \mathrm{T} \tag{15}$$

The equation 15 depicts the benchmark (AR-1) model in which current GDP_t is forecasted using the previous period's GDP_{t-1} values. This study also used models that included only consumer confidence and Google Trends data. These model specifications allowed a thorough examination of the soft data sets.

$$GDP_t = \beta_0 + \beta_1 Confidence_t + \varepsilon_t \qquad t = 1, \dots, \mathrm{T} \tag{16}$$

The confidence model is in equation 16, where the current GDP is nowcasted with only country's consumer confidence data.

$$GDP_t = \beta_0 + \beta_1 Google_{it} + \varepsilon_t \qquad \begin{array}{l} t = 1, \dots, \mathrm{T} \\ i = 1, \dots, \mathrm{N} \end{array} \tag{17}$$

The model that included only the Google Trends data is in equation 17. In it, the current GDP is nowcasted using the current Google Trends data. This thesis uses equation 17 for three different levels of Google Trends data. Firstly, this thesis uses it for the entire Google Trends data. Secondly, this thesis utilizes equation 17 for broad category models, i.e. models that were constructed using the dimension reduction methods[3].

[3] This thesis was unable to eliminate all of the future GDP influence in the partial least squares (PLS) dimension reduction method. In other words, partial least squares models were created in pseudo-out-of-sample. However, forecasts use Google data that have been constructed with respect to ex-post GDP data. This construction can give the PLS method forecasts an advantage in GDP forecasting.

Thirdly and finally, this master's thesis applies equation 17 for Google subcategory models that this thesis formed with the LASSO shrinkage method. It is noteworthy that this thesis used the LASSO method for ex-post data, i.e. data available in 2018. However, the optimal category models, which were selected by the LASSO, were estimated in the manner of pseudo-out-of-sample exercise.

(18) $$GDP_t = \beta_0 + \beta_1 GDP_{t-1} + \beta_2 Confidence_t + \varepsilon_t \qquad t = 1, \ldots, \mathrm{T}$$

The equation 18 depicts model where current GDP_t is nowcasted using the previous GDP values and the current consumer confidence data $Confidence_t$. This kind of nowcasting is possible because is GDP quarterly data and consumer confidence is –more frequent– monthly data.

(19) $$GDP_t = \beta_0 + \beta_1 GDP_{t-1} + \beta_2 Google_{it} + \varepsilon_t \qquad \begin{matrix} t = 1, \ldots, \mathrm{T} \\ i = 1, \ldots, \mathrm{N} \end{matrix}$$

Equation 19 is similar to the earlier equation, but now the Google Trends data is used as a nowcasting factor for the current GDP.

(20) $$GDP_t = \beta_0 + \beta_1 GDP_{t-1} + \beta_2 Confidence_t + \beta_3 Google_{it} + \varepsilon_t$$

$$\begin{matrix} t = 1, \ldots, \mathrm{T} \\ i = 1, \ldots, \mathrm{N} \end{matrix}$$

The equation 20 depicts nowcasting model that includes both $Google_{it}$ $Confidence_t$ variables. This sort of combined model specification allows for an additional comparison between the Google Trends and consumer confidence data.

(21) $$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(Predicted\ values_t - Actual\ values_t)^2}{T}} \qquad t = 1, \ldots, T$$

This thesis compared all model forecasts by their *Root mean squared errors* (RMSE), which is in equation 21. RMSE have been relatively popular statistic to evaluate model's accuracy and many previously mentioned papers have applied it, e.g. Bańbura et. al (2013), McLaren and Shanbhogue (2011), Bańbura and Rünstler (2011).

In RMSE, predicted values are being ex-ante compared to the actual values. In other words, RMSE compares forecasted GDP values to the actual realized GDP values. This comparison is increased to the power of two, to assess the possible negative values. The squaring also emphasis large errors over small ones. These values are then summed together to cover every observation.

Finally, by dividing it by the amount of the observations, one can get the average RMSE result. The RMSE is useful when comparing different models' forecasting performance. Furthermore, the model, which has the lowest RMSE score, is the most accurate. However, by itself, the statistic does not provide any useful information.

5 RESULTS AND ANALYSIS

This chapter 5 presents models' results, which include models' RMSE scores, figures and model estimates. The section begins by discussing results for the benchmark and confidence models. After that, the Google models' results are presented in three different subsections depending on the levels of data.

What is more, this section includes a subsection for cross-validation method, which analyses nowcasting models' robustness. The following paragraph examines the previously found relation with Google searches and policy-related uncertainty. Chapter 5 concludes with the discussion of the results, i.e. how this thesis's findings relate to earlier studies and are the results reliable.

5.1.1 Benchmark and consumer confidence results

Results in the following tables are divided into two different segments depending on the data format. "Three-month average" relates to data that this master's thesis formed by averaging every three months. The "every third-month" data represents the most recent data available because the GDP is a quarterly statistic.

The numbers in parenthesis represent the previously presented model equations. On top of that, this study calculated all of the RMSE scores with the formula shown in equation 21. The following table 4 presents results for both Finland's benchmark and confidence models.

Table 4: RMSE results of Finland's models (15), (16) and (18)

Country:	**Finland**
Model:	**RMSE Results:**
AR-1 (15)	1.775

Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Confidence (16)	1.316	1.300
AR-1 and Confidence (18)	1.767	1.670

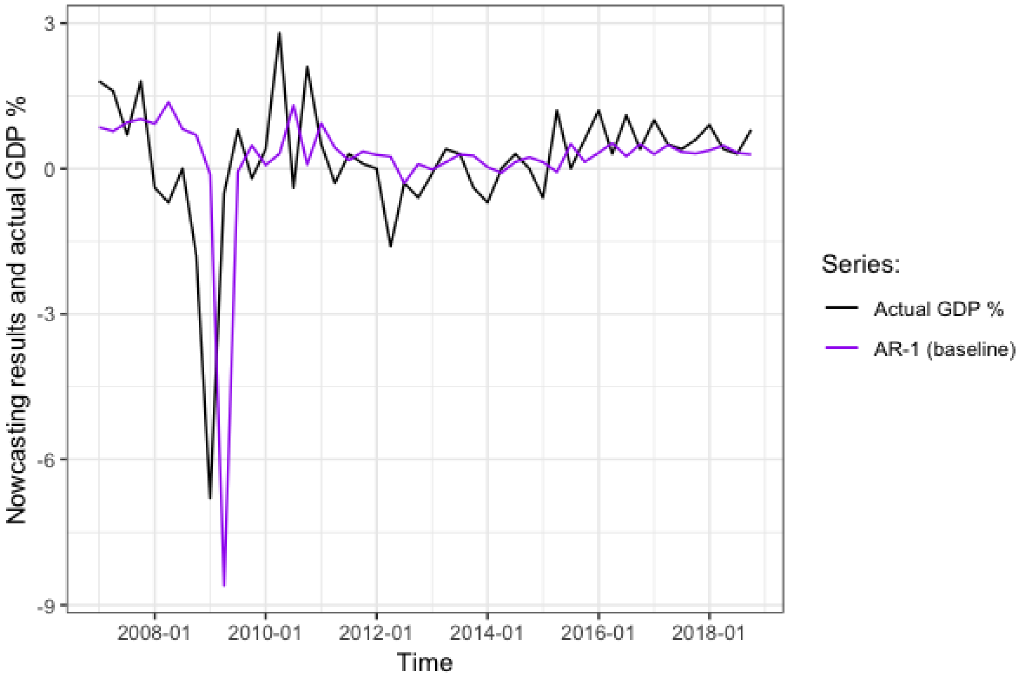


Figure 9: Finland's benchmark (AR-1) model and actual GDP growth

Table 4 and figure 9 imply that Finland's previous GDP changes can generate somewhat useful information about the current GDP. However, this information is useless in the most volatile times. For example, during the 2008 financial crises, the benchmark model with AR-1 lag variable generated GDP forecasts with a more profound and longer-lasting recession.

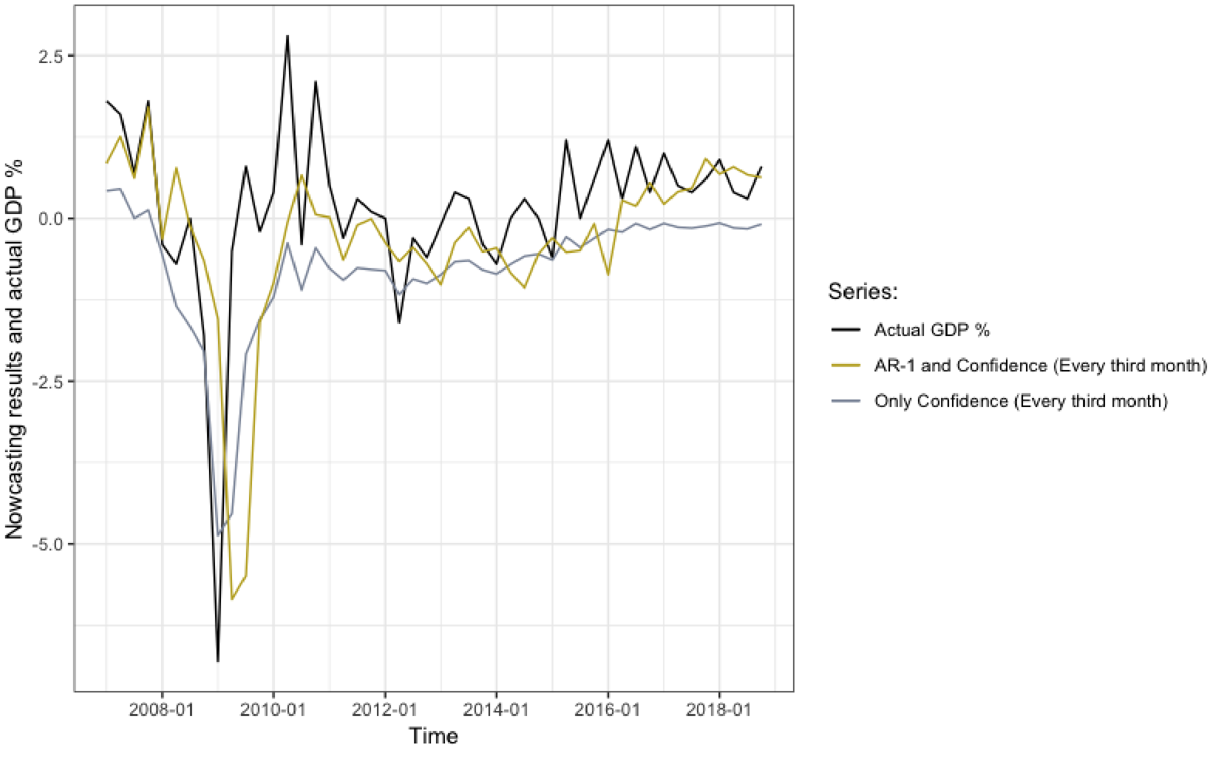


Figure 10: Finland's consumer confidence models and actual GDP growth

According to RMSE results in table 4, consumer confidence has an impressive ability to nowcast Finland's current GDP. Furthermore, figure 10 illustrates models that included the consumer confidence variable. Figure 10 also further suggests that a model that included only the consumer confidence were able to follow Finland's GDP growth carefully.

Moreover, amidst the 2008 financial crises, the fall in consumer confidence happened simultaneously with Finland's GDP. As the recession deepened people's confidence in their economy declined. These results confirm survey data's remarkable nowcasting abilities, which were previously found by Bańbura & Rünstler (2011) and Giannone et al. (2008). The next table 5 and figure 11 present Germany's benchmark and confidence model's results.

Table 5: RMSE results of Germany's models (15), (16) and (18)

Country:	**Germany**
Model:	**RMSE Results:**
AR-1 (15)	1.218

Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Confidence (16)	1.032	1.028
AR-1 and Confidence (18)	1.273	1.280

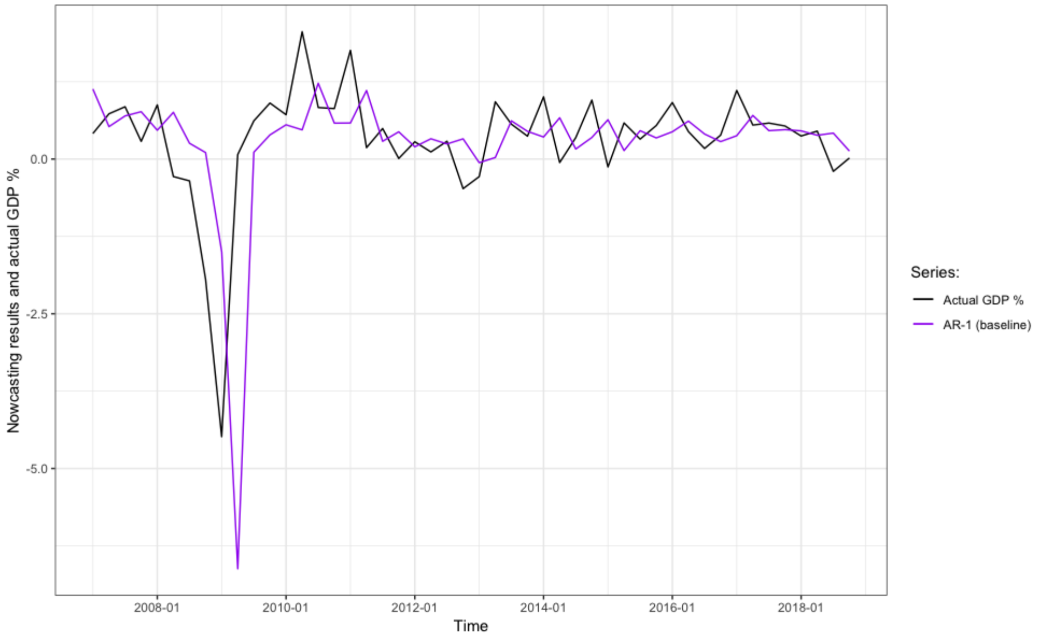


Figure 11: Germany's benchmark (AR-1) model and actual GDP growth

As seen in table 5 and figure 11, Germany's previous GDP values can generate slightly useful information regarding the current GDP. Similarly, to Finland's benchmark model, this information is rather useless in the most volatile times. However, Germany's similarities with Finland end here as the consumer confidence model's results differ considerably. These differences are visible in the following figure 12.

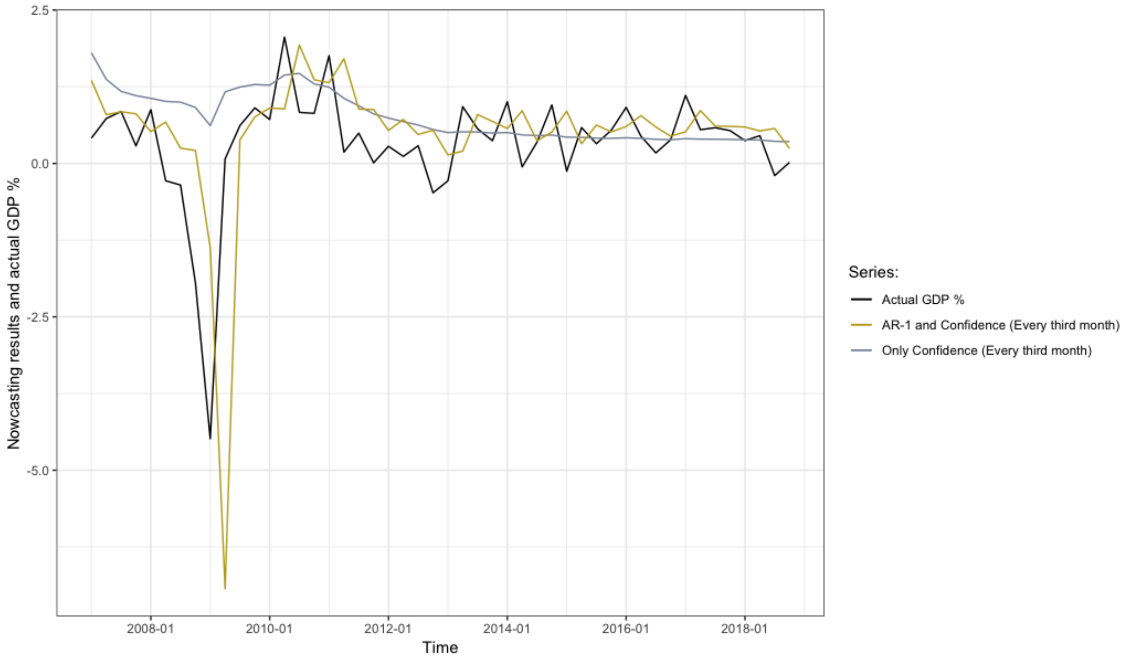


Figure 12: Germany's consumer confidence models and actual GDP growth

Table 5 presents the univariate confidence model's RMSE results, which are much lower than the benchmark model. Despite this, figure 12 suggest that Germany's consumer confidence model is not that accurate in nowcasting the current GDP. What is more, the confidence models' estimates are not sufficiently varied, which make their practical usage rather tricky. Thus, with these specifications, Germany's confidence model produces poor nowcasting results.

5.1.2 Initial Google Trends results

Now, it is appropriate to discuss one of the most central topics of this master's thesis, i.e. Google models' nowcasting performance. Google Trends data's high dimensionality was reduced using principal component analysis (PCA) and partial least squares (PLS). This thesis applied these methods both before and during the nowcasting exercise.

Appendix 4 shows results for models, which used dimension reduction methods before the exercise. Furthermore, this thesis applied these methods for different levels of Google Trends data. The following section reports RMSE scores and nowcasting estimates for models that used the entire Google data. The consecutive sections report results for more specific levels, i.e. Google Trends data's broad and subcategories. Also, these RMSE results are in four different segments depending on the dimension reduction method and the data.

The first Google models RMSE scores are in table 6 and 7. These models included a specific country's entire Google Trends data, i.e. all the 181 initial subcategories from Finland and Germany. Strictly speaking, these models were examining how the quantity of Google searches are related to a country's GDP growth. Could an increase in Google searches be an indicator of increased economic activity? The next analysis discusses first Finland's Google models and then Germany's models.

Table 6: Models that included entire Finland's Google Trends data

Country:	**Finland**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google (17)	1.445	1.455
AR-1 and Google (19)	1.635	1.692
AR-1 with Google and Confidence (20)	1.699	1.665

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google (17)	1.509	1.503
AR-1 and Google (19)	1.680	1.713
AR-1 with Google and Confidence (20)	1.717	1.692

Table 6 results state that regardless of the dimension reduction method, the univariate models were the most accurate. In addition, principal component (PCA) method produced the most accurate model with RMSE score of 1.445. RMSE results are also quite similar between the different data formations, i.e. three-month average and every third month.

In conclusion, it appears that Google Trends data is somewhat related to Finland's GDP growth. This relation is because RMSE scores are not that far from consumer confidence models. Google Trends models can also generate lower RMSE scores than the benchmark (AR-1) model. Subsequent figures 13 and 14 show Google models nowcasting estimates against Finland's GDP and consumer confidence.

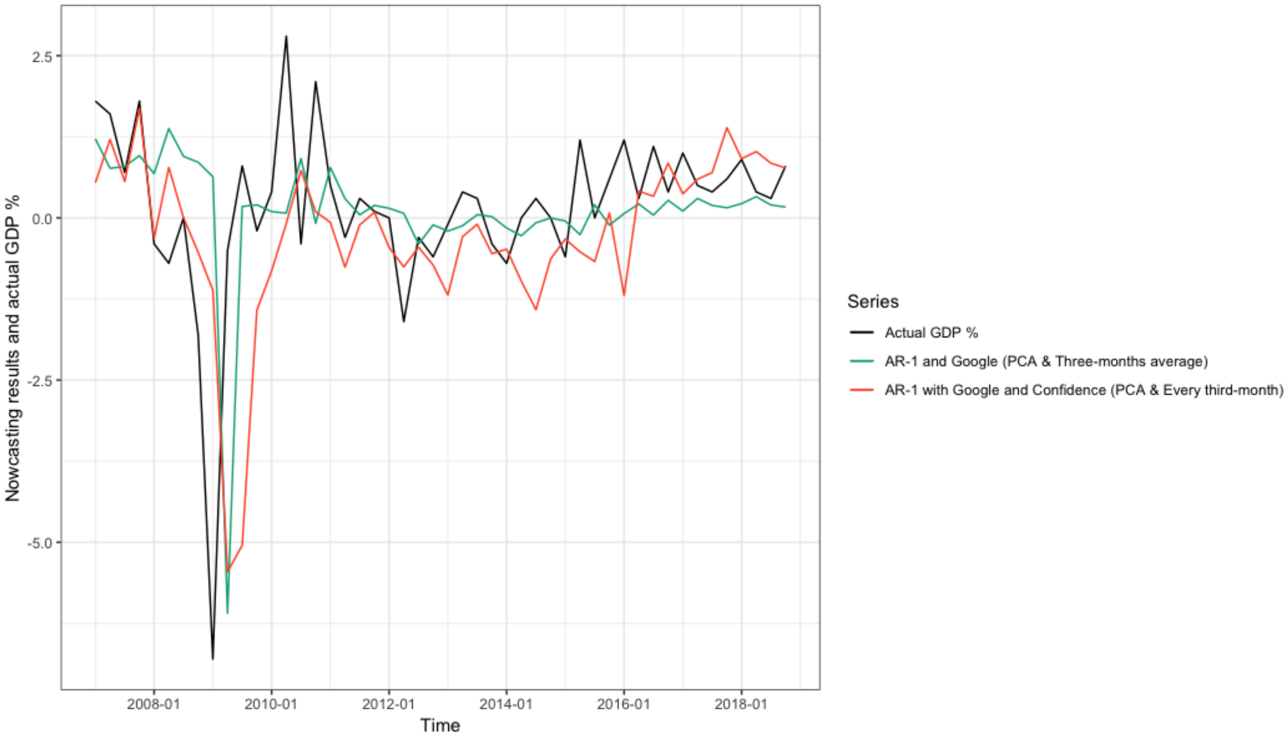


Figure 13: Google models and Finland's GDP growth

Figure 13 portrays Google models that included AR-1 lag and consumer confidence variables. It appears as that these models are producing lagged estimates, as one would expect. However, it seems that model with confidence variable is generating more volatile nowcasting estimates. This higher estimate volatility is evident in the following figure 14.

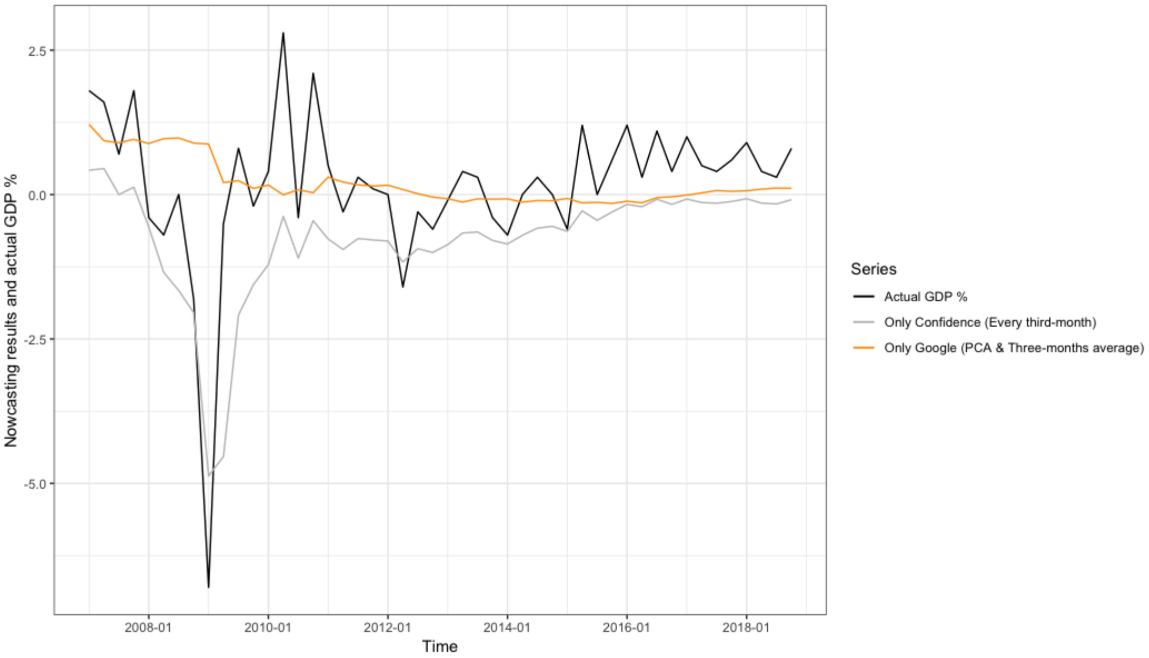


Figure 14: Leading Google model and Finland's GDP growth

Figure 14 presents the leading Google model against the leading consumer confidence model. Both Google and consumer confidence models can achieve similar results, as both had the same reaction to the 2008 financial crisis. Consumer confidence decreased because people were anxious about the recession. Google searches also appeared to fall, which may be caused by the fact that people were planning to consume fewer durable goods as most of the categories relate to consumption.

Table 7: Models that included entire Germany's Google Trends data

Country:	**Germany**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google (17)	0.983	0.980
AR-1 and Google (19)	1.138	1.157
AR-1 with Google and Confidence (20)	1.177	1.231

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google (17)	0.906	0.906
AR-1 and Google (19)	1.100	1.121
AR-1 with Google and Confidence (20)	1.401	1.405

Germany's univariate Google models, found in table 7, produced the lowest RMSE results. According to the RMSE scores, the most accurate Google model was constructed using the partial least squares (PLS). However, these results did not seem to depend on the data formation.

It is also interesting that the inclusion of consumer confidence data increased RMSE scores. In other words, consumer confidence data weakened Google model's accuracy. Overall, the Google model's RMSE results strengthen the view that Germany's Google searches are related to its current GDP growth. Following figures, 15 and 16 further discuss this relationship.

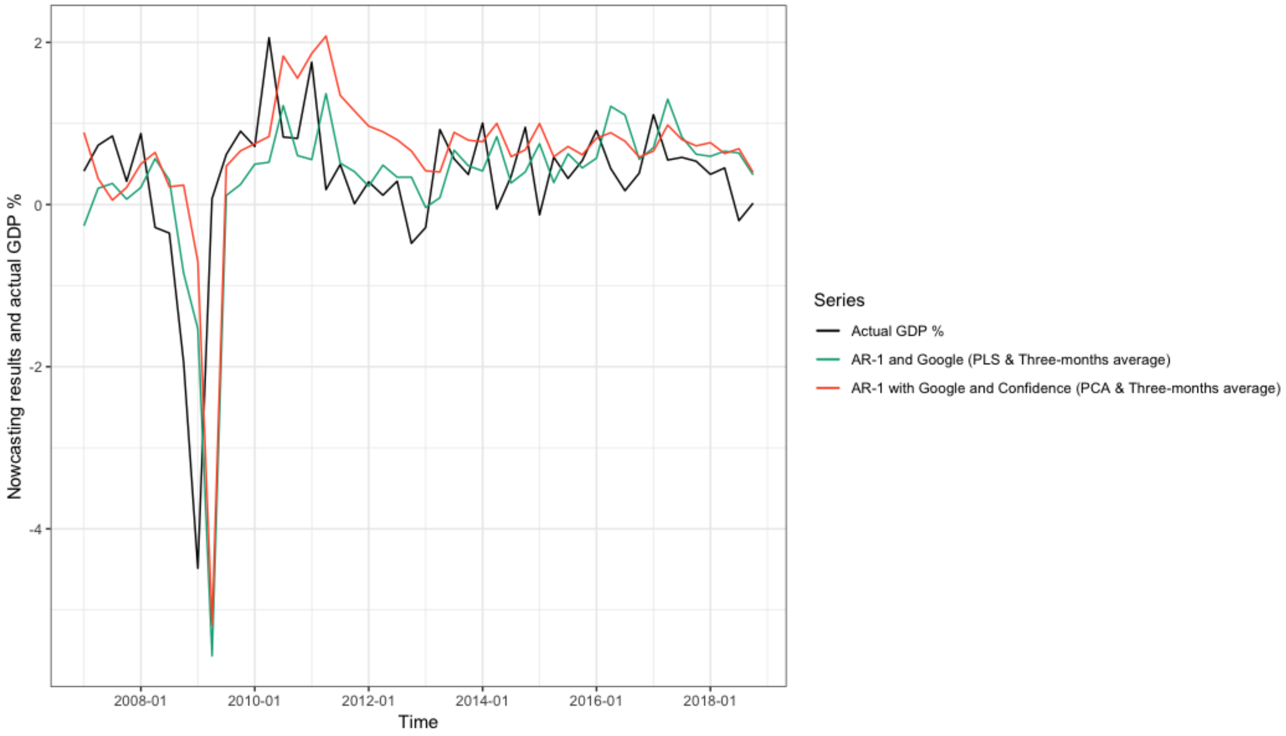


Figure 15: Google models and Germany's GDP growth

Figure 15 displays Germany's Google models, which included AR-1 lag and consumer confidence. Because of the lag variable, both models generated lagged nowcasting estimates. However, compared to Finland, both of these model estimates are quite similar. Moreover, there are only a few differences between them. This similarity might be because the German consumer confidence is not able to generate any further information to the model.

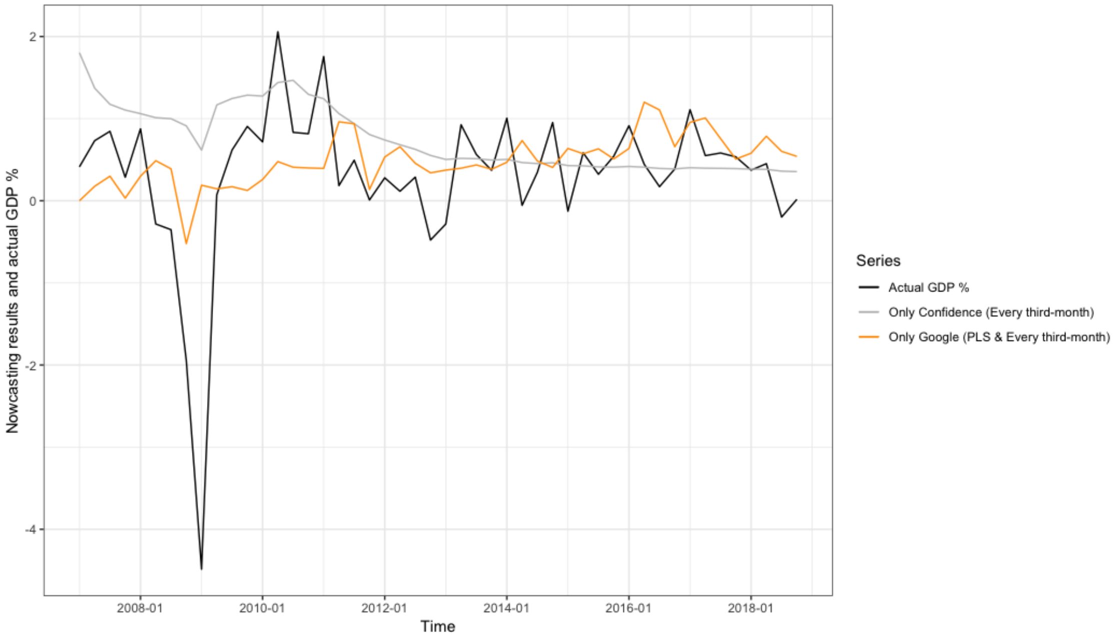


Figure 16: Leading Google model and Germany's GDP growth

Figure 16 portrays both Germany's consumer confidence model and the Google model. Unlike in Finland, Germany's Google model produces significantly more variation. Moreover, it seems that Germany's consumer confidence hardly decreased in the 2008 financial crises. However, Google searches decreased significantly, and it happened amidst the crises. This decrease could be because financial crises affected almost immediately people's consumption habits.

Both Finland's and Germany's results confirm the Vosen and Schmidt (2011, 573–576) finding that Google searches relating to consumption decreased in amidst of the 2008 financial crisis. However, the Google model initial results are somewhat mixed. In Finland, the Google model produced more stable nowcasting estimates than consumer confidence model. On the contrary, Germany's Google models produced more volatile results.

5.1.3 Google category analysis and results

Despite these initial Google results, it is still unclear, what are the underlying factors in these searches. Therefore, the category analysis is examining, are there any specific search terms categories that are relating to a country's GDP growth. This category analysis has a two-part structure.

The first part examines the 16 different Google Trends broad categories. This master's thesis created these broad categories during the nowcasting exercise using the dimension reduction method, and they are visible in the earlier table 2. In addition, this thesis constructed categories before the exercise, and their results are in appendix 4. These models using pre-exercise categories gave more emphasis on the partial least squares (PLS) method, but overall, they produced a somewhat similar result.

The second part of the category analysis focused on the 181 different Google Trend subcategories that are in appendix 1. The subcategory analysis was performed using the LASSO shrinkage method, which selected the optimal search categories. Both category analyses used similar models shown in section 4. Following tables 8 and 9 presents RMSE results for the broad category models. In other words, the model results for the univariate Google model in equation (17). Rest of the broad category results are in appendix 3.

Table 8: RMSE results of Finland's Google category models (17)

Country:	**Finland**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google categories (17)		
Autos & Vehicles	1.443	1.427
Beauty & Fitness	1.475	1.488
Business & Industrial	1.441	1.388
Computers & Electronics	1.432	1.434
Food & Drink	1.443	1.457
Health	1.450	1.399
Home & Garden	1.407	1.418
Internet	1.461	1.448
Investing	1.461	1.448
Jobs	1.376	1.404
Law	1.458	1.442
News	1.350	1.378
Real Estate	1.372	1.382
Shopping	1.465	1.453
Sports	1.440	1.423
Travel	1.431	1.428

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google categories (17)		
Autos & Vehicles	1.430	1.468
Beauty & Fitness	1.624	1.579
Business & Industrial	1.697	1.479
Computers & Electronics	1.440	1.436
Food & Drink	1.447	1.427
Health	1.609	1.530
Home & Garden	1.458	1.457
Internet	1.510	1.499
Investing	1.421	1.419
Jobs	1.584	1.468
Law	1.483	1.458
News	1.415	1.417
Real Estate	1.483	1.466
Shopping	1.500	1.444
Sports	1.460	1.462
Travel	1.445	1.435

Table 8 presents the results for the 16 different broad categories. It appears that in Finland, all of the Google Trend broad category models have lower RMSE results than the benchmark AR-1 model, i.e. they are more accurate. The most precise broad categories being Jobs, Real Estate and News. All of which were constructed using the principal component analysis (PCA) method and three-month average data. The following figure 17 plots two of these leading nowcasting models against Finland's GDP growth.

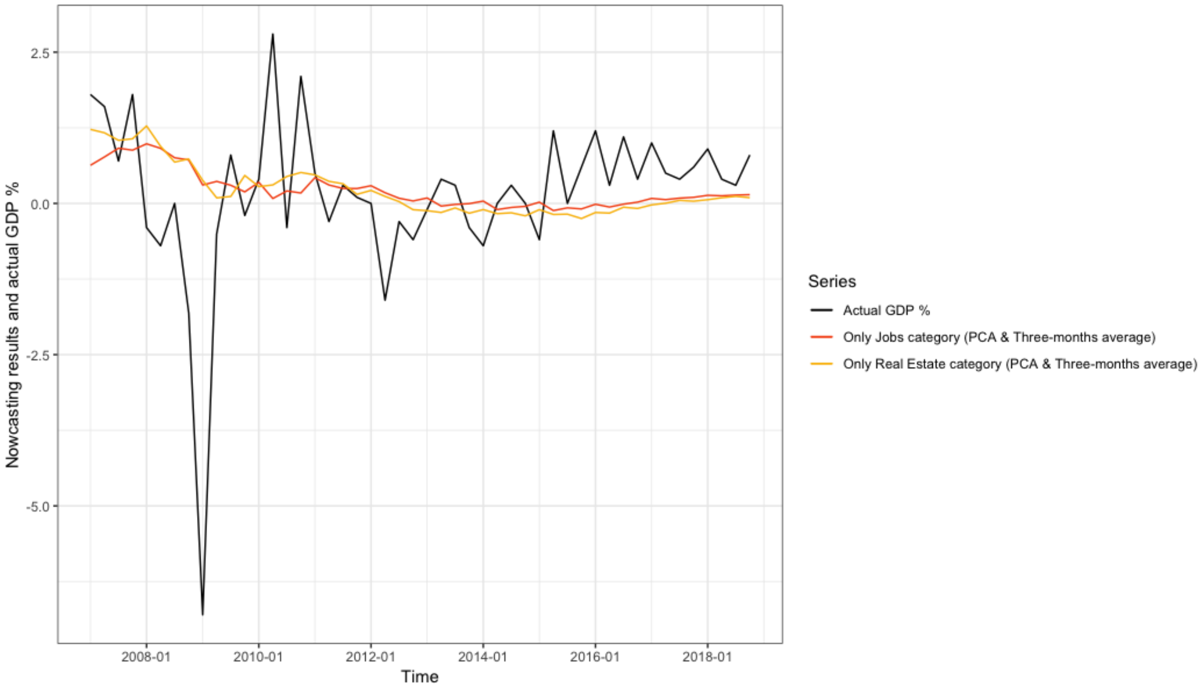


Figure 17: Two of the leading Google category models and Finland's GDP growth

Figure 17 depicts two of the leading Google Trends category models against Finland's GDP growth. It also confirms the earlier result that Google Trends nowcasts relatively small changes to current GDP. Additionally, Google Trends categories did not seem to have a significant reaction to the 2008 crises. However, figure 17 states that Google searches regarding jobs and real estate did have a minor decrease after the financial crises. It is possible to break down the factors of these broad category estimates by reviewing their subcategories in appendix 1.

It may be that financial crises affected people's Google searches for Real Estate related search terms. Namely, fewer people could have been searching for new housing and mortgages. It is also interesting that searches for jobs had decreased in the aftermath of financial crises. This decrease contradicts Tuhkuri (2014) results, which found that search terms related to unemployment increased after the financial crises. This different result may be because Tuhkuri (2014) used six different keywords to proxy Finland's unemployment. This thesis used Googles own categories to intermediate unemployment. Thus, it could be that Google is not able to categorise Finland's unemployment search terms correctly.

Moreover, when analysing the Googles broad categories nowcasting accuracy, all the RMSE results are higher than consumer confidence models. In other words, consumer confidence outperformed all of the Google Trends broad category models. This conclusion is characterised by the following figure 18.

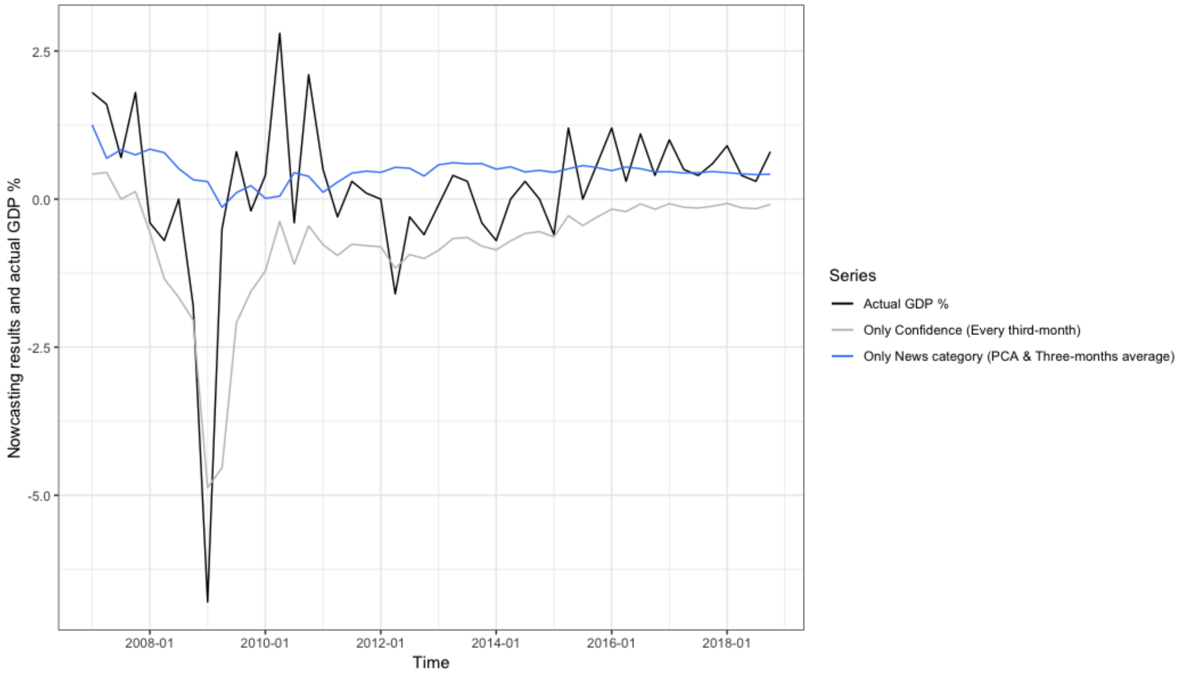


Figure 18: Confidence and the leading Google model against Finland's GDP growth

Figure 18 portrays how Finland's leading broad category, i.e. the News model, fares against the consumer confidence model. This thesis formed the leading News model with the principal component method (PCA). What is more, the 2008 downturn seemed to have little effect on people's searches for news in Finland. However, as before consumer confidence model can foreshadow the 2008 financial crises. News category model's reaction is only ex-post, at best.

Besides, consumer confidence models nowcasts are overall more in line with the actual GDP growth. According to figure 18, when Finland's GDP has surged, most notably in 2011, the consumer confidence models growth estimates increased. Thus, consumer confidence models generated the most accurate and reliable nowcasting estimates when models included only one variable. The next analysis examines Germany's broad category models.

Table 9: RMSE results of Germany's Google category models (17)

Country:	**Germany**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google categories (17)		
Autos & Vehicles	0.977	0.981
Beauty & Fitness	0.994	0.991
Business & Industrial	0.974	0.974
Computers & Electronics	0.967	0.967
Food & Drink	1.008	1.002
Health	0.981	0.981
Home & Garden	0.983	0.996
Internet	0.966	0.965
Investing	1.005	0.994
Jobs	0.972	0.974
Law	0.979	0.983
News	0.947	0.986
Real Estate	0.929	0.991
Shopping	0.984	0.973
Sports	0.971	0.977
Travel	0.977	0.981

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google categories (17)		
Autos & Vehicles	0.919	0.914
Beauty & Fitness	0.974	1.020
Business & Industrial	1.013	0.984
Computers & Electronics	1.186	1.053
Food & Drink	0.979	0.961
Health	0.981	1.379
Home & Garden	0.947	0.959
Internet	1.073	1.023
Investing	1.014	0.980
Jobs	1.012	0.935
Law	1.508	1.556
News	1.004	0.888
Real Estate	0.954	0.985
Shopping	1.014	0.954
Sports	0.940	0.934
Travel	0.948	0.937

Similar to earlier table 8, table 9 presents Germany's Google model results for 16 different broad categories. Somewhat like in Finland, almost all these Google models produced lower RMSE results than the benchmark AR-1 model. Only Law broad category, which was created by PLS generated higher RMSE results than the benchmark model. Thus, almost of all this thesis's univariate Google models were able to outperform their benchmark models.

In Germany, the most accurate broad category models were Autos & Vehicles, Real Estate and News. These models did not have any superior dimension reduction method, and the leading models were constructed with different methods. The following figures, 19 and 20, depict these leading category models against Germany's GDP growth.

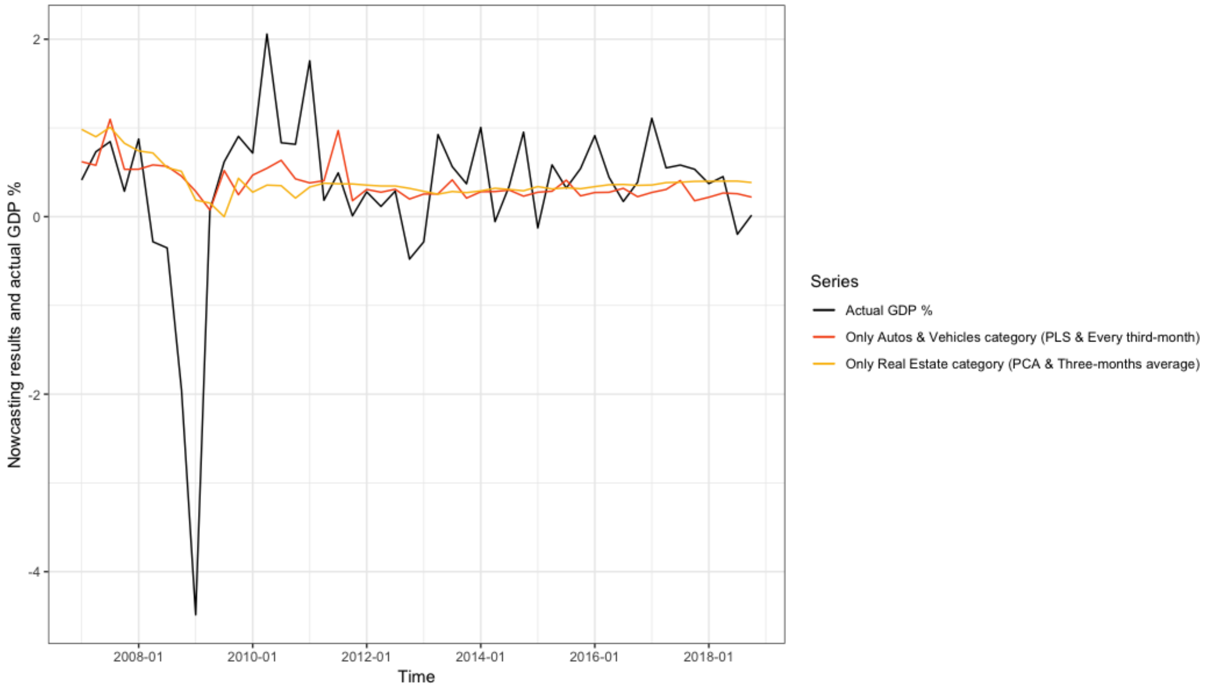


Figure 19: Two of the leading Google category models and Germany's GDP growth

Figure 19 presents the Google model's results for Autos & Vehicles and Real Estate categories. Both of these models produced quite smooth estimates. Both of the category models estimates decreased shortly after the financial crises. This decrease might be because, after the crises, the people were using less of their time to search for new housing and cars.

After the crises, Autos & Vehicles category nowcasted significant increase to Germany's GDP. This increase was in line with the actual GDP growth. This result could suggest that Google searches related to Autos & Vehicles have some relationship with Germany's GDP changes. This relation could be because the automotive industry is a large part of Germany's manufacturing sector. This is somewhat in line with Götz and Knetsch (2019) finding of Google information relating to the manufacturing industry in Germany.

Still, Real Estate model's estimates are too smooth for practical purposes, and they do not seem to follow Germany's GDP growth carefully. Nevertheless, the following figure 20 presents the leading broad category Google model in Germany.

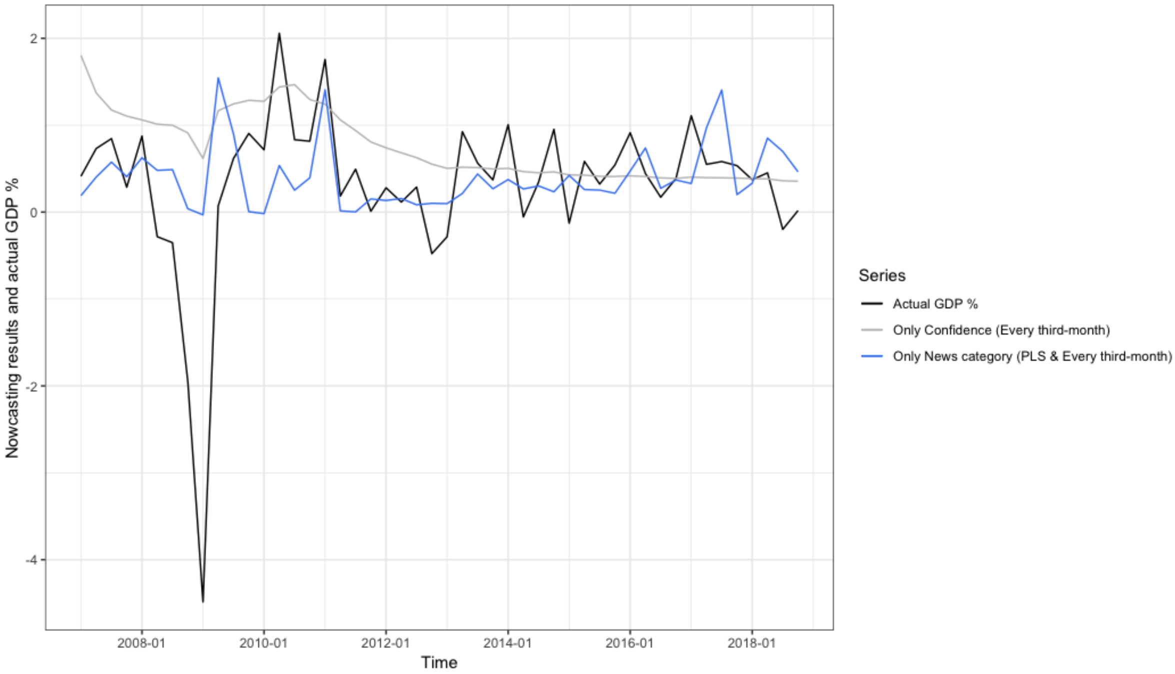


Figure 20: Confidence and the leading Google model against Germany's GDP growth

Figure 20 depicts nowcasting estimates for both confidence and News category model. As previously stated, Germany's confidence model produces rather stable forecasts. However, the leading broad category model generates quite intriguing results. News category estimates increased significantly after the financial crises. This increase might be because after the crises actualized, the people were searching for news about the financial crises, i.e. people were doing a considerable amount of Google searches when macroeconomic, and policy-related uncertainty was high.

This result is similar to Donadelli (2015) finding that in Google searches have a positive relationship with policy-related uncertainty. However, after the crises, News categories relationship with GDP changed. Figure 20 also suggests that post-crisis News category had a mostly positive relation with Germany's GDP growth.

So far, this master's thesis has mainly discussed category analysis's univariate results. Google multivariate models both in Finland and Germany generated significantly inferior RMSE results, i.e. nowcasting equations 19 and 20. These multivariate results are in appendix 3. According to RMSE scores, the most accurate Finnish multivariate Google models generally contain both Google categories and AR-1 variable. In other words, they did not contain consumer confidence variable. Thus, it appears that the Google category and AR-1 variable can capture most of the relevant information regarding Finland's GDP.

Similar to Finland, the inclusion of consumer confidence data also weakened Germany's multivariate models. Therefore, univariate and multivariate results suggest that Google data is capable of capturing GDP information more effectively than consumer confidence in Germany at least. For a clearer picture, the following tables 10 & 11 presents five leading nowcasting models and their RMSE results and estimates.

Table 10: Model estimates for five leading models in Finland

Country	**Finland**			
RMSE result:	1.300			
Model:	**Every third-month**			
Only Confidence (16)	Estimate	Std. Error	t-value	p-value
Coefficients:				
Intercept	-0.207	0.232	-0.892	0.376
Confidence	0.147	0.050	2.960	**0.005***

p <.0001 ’***’, p <.001 ’***’, p <.01 ’**’, p <0.05 ’*’

RMSE result:	1.316			
Model:	**Three-months average**			
Only Confidence (16)	Estimate	Std. Error	t-value	p-value
Coefficients:				
Intercept	-0.164	0.247	-0.665	0.509
Confidence	0.133	0.054	2.459	**0.017***

p <.0001 ’***’, p <.001 ’***’, p <.01 ’**’, p <0.05 ’*’

Dimension reduction method:	**Principal component analysis (PCA)**			
RMSE result:	1.350			
Model:	**Three-months average**			
Only Google categories (17)	Estimate	Std. Error	t-value	p-value
Coefficients:				
Intercept	0.295	0.170	1.737	0.088
News	0.039	0.066	0.590	0.558

p <.0001 ’***’, p <.001 ’***’, p <.01 ’**’, p <0.05 ’*’

Dimension reduction method:	**Principal component analysis (PCA)**			
RMSE result:	1.372			
Model:	**Three-months average**			
Only Google categories (17)	Estimate	Std. Error	t-value	p-value
Coefficients:				
Intercept	0.295	0.168	1.752	0.085
Real Estate	0.077	0.068	1.137	0.260

p <.0001 ’***’, p <.001 ’***’, p <.01 ’**’, p <0.05 ’*’

Dimension reduction method:	**Principal component analysis (PCA)**			
RMSE result:	1.376			
Model:	**Three-months average**			
Only Google categories (17)	Estimate	Std. Error	t-value	p-value
Coefficients:				
Intercept	0.295	0.169	1.744	0.087
Jobs	-0.108	0.121	-0.886	0.379

p <.0001 ’***’, p <.001 ’***’, p <.01 ’**’, p <0.05 ’*’

As found in table 10, the most accurate models in Finland were the consumer confidence models, which also had significant coefficient estimates. The most accurate Google models were all constructed using the principal component analysis (PCA) method. These Google models included categories relating to News, Real Estate and Jobs.

News category had the lowest RMSE score; hence, it is the most accurate Google broad category model to nowcast Finland's GDP. However, the News category model's estimate is not nearly significant, with a p-value of 0.558. In summary, broad category analysis, suggests that consumer confidence is the most relevant data source to nowcast Finland's GDP growth. The following table 11 presents estimates for the five leading nowcasting models in Germany.

Table 11: Model estimates for five leading models in Germany

Country:	**Germany**			
Dimension reduction method:	**Partial least squares (PLS)**			
RMSE result:	0.888			
Model:	**Every third-month**			
Only Google categories (17)	Estimate	Std. Error	t-value	p-value
Coefficients:				
Intercept	0.379	0.111	3.411	0.001**
News	0.028	0.011	2.514	0.015*

p <.0001 ’***’, p <.001 ’***’, p <.01 ’**’, p <0.05 ’*’

Dimension reduction method:	**Partial least squares (PLS)**			
RMSE result:	0.914			
Model:	**Every third-month**			
Only Google categories (17)	Estimate	Std. Error	t-value	p-value
Coefficients:				
Intercept	0.379	0.113	3.340	0.002**
Autos & Vehicles	0.010	0.005	1.934	0.058

p <.0001 ’***’, p <.001 ’***’, p <.01 ’**’, p <0.05 ’*’

Dimension reduction method:	**Partial least squares (PLS)**			
RMSE result:	0.919			
Model:	**Three-months average**			
Only Google categories (17)	Estimate	Std. Error	t-value	p-value
Coefficients:				
Intercept	0.379	0.115	3.302	0.002*
Autos & Vehicles	0.007	0.004	1.547	0.127

p <.0001 ’***’, p <.001 ’***’, p <.01 ’**’, p <0.05 ’*’

Dimension reduction method:	**Principal component analysis (PCA)**			
RMSE result:	0.929			
Model:	**Three-months average**			
Only Google categories (17)	Estimate	Std. Error	t-value	p-value
Coefficients:				
Intercept	0.379	0.117	3.234	0.002*
Real Estate	0.002	0.043	0.049	0.961

p <.0001 ’***’, p <.001 ’***’, p <.01 ’**’, p <0.05 ’*’

Dimension reduction method:	**Partial least squares (PLS)**			
RMSE result:	0.934			
Model:	**Every third-month**			
Only Google categories (17)	Estimate	Std. Error	t-value	p-value
Coefficients:				
Intercept	0.379	0.113	3.351	0.001*
Sports	0.011	0.005	2.037	0.046*

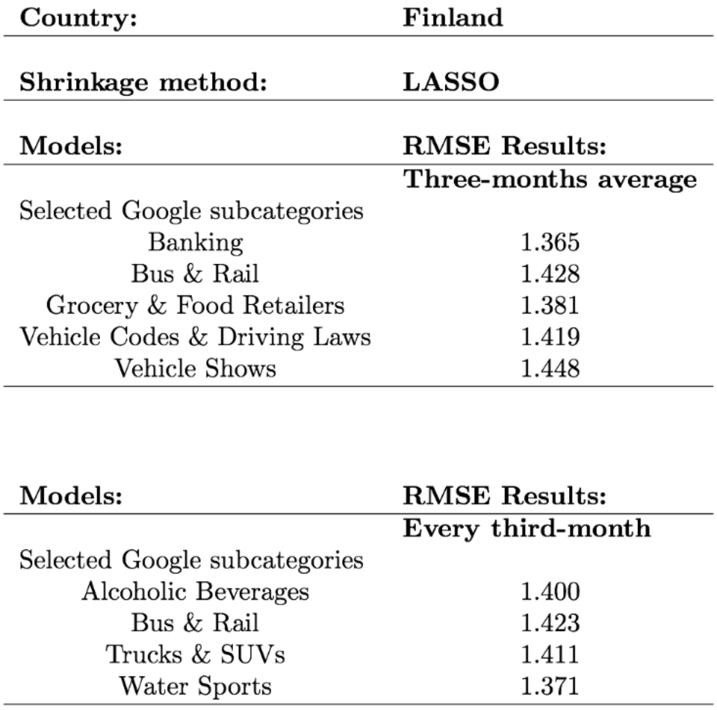
p <.0001 ’***’, p <.001 ’***’, p <.01 ’**’, p <0.05 ’*’

As seen from the table 11, the five leading nowcasting models in Germany were all Google models. The five leading Google broad category models were News, Autos & Vehicles, Real Estate and Sports. Unlike Finland, most of these were constructed using the partial least squares (PLS) method. It also noteworthy that two of the five Google models had significant coefficients for the Google variables.

Broad category analysis suggests that in Finland consumer confidence model was consistently the most accurate and robust nowcasting model. In Germany, consumer confidence falls behind, and the most accurate model was the News category model. However, it is still unclear, what is the driving force behind both Finland and Germany's broad categories models. In other words, what are the primary Google subcategories affecting the leading broad categories?

Following part of the analysis is examining the Google Trends subcategories that are in appendix 1. Especially, is there a Google Trends subcategory that has an especially close relationship with a country's GDP growth? These optimal subcategories were selected using LASSO shrinkage method. In addition, the LASSO method was applied separately for the three-month average data and every third-month data. This master's thesis constructed subcategory models based on the univariate Google models, i.e. equation 17. Subcategory results are in tables 12 & 13.

Table 12: Finland's subcategory models selected by LASSO shrinkage method



Country:	**Finland**
Shrinkage method:	**LASSO**
Models:	**RMSE Results:**
	Three-months average
Selected Google subcategories	
Banking	1.365
Bus & Rail	1.428
Grocery & Food Retailers	1.381
Vehicle Codes & Driving Laws	1.419
Vehicle Shows	1.448

Models:	**RMSE Results:**
	Every third-month
Selected Google subcategories	
Alcoholic Beverages	1.400
Bus & Rail	1.423
Trucks & SUVs	1.411
Water Sports	1.371

According to table 12, even with optimal Google subcategories, Finland's consumer confidence is still able to dominate the comparison. In other words, previous table 4 revealed consumer confidence model's RMSE score to be 1.300. Thus, it seems that Google Trends is consistently secondary regards to consumer confidence data in nowcasting.

Even with different levels of Google Trends data, the consumer confidence can generate the lowest RMSE score, i.e. the most accurate nowcasting estimates. In Finland, the most accurate subcategory model included search terms related to banking. Thus, the Banking subcategory was the driving force behind the Investment category with RMSE score of 1,365. The following figure 21 depicts these estimates against Finland's GDP growth.

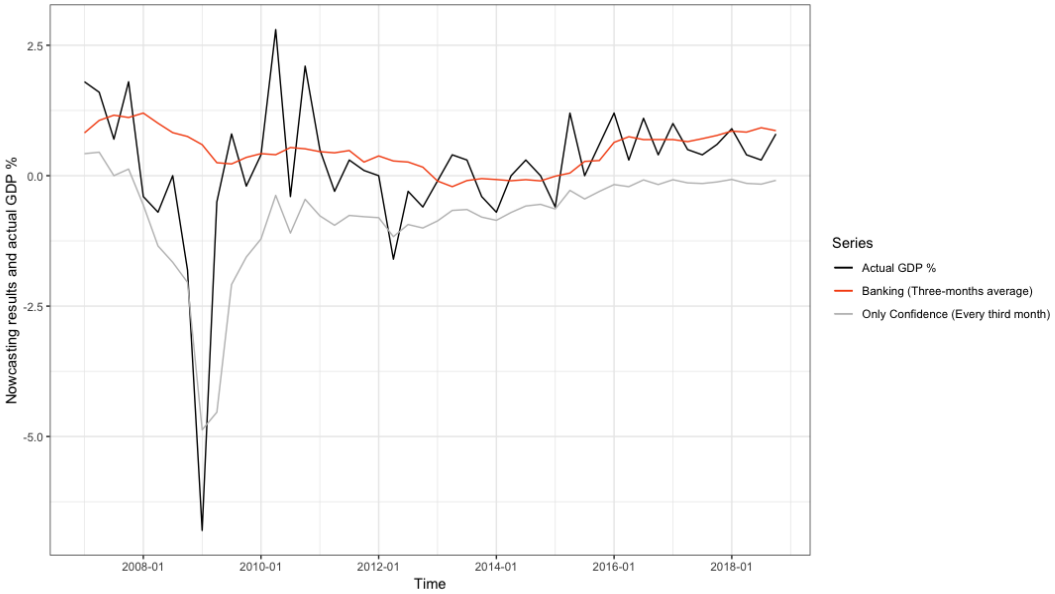
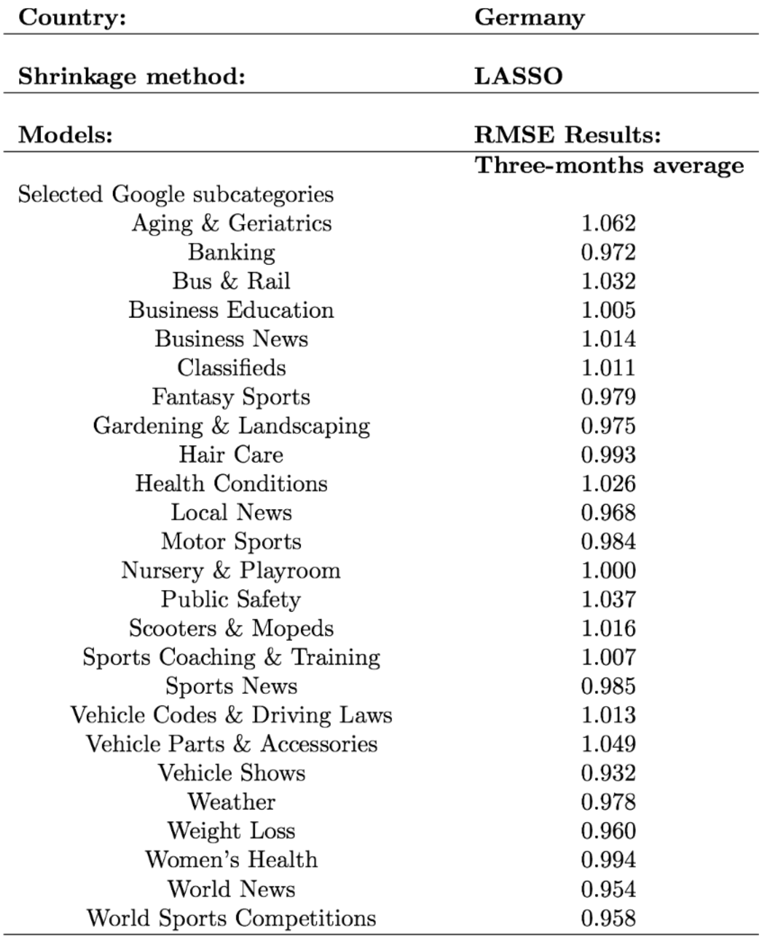


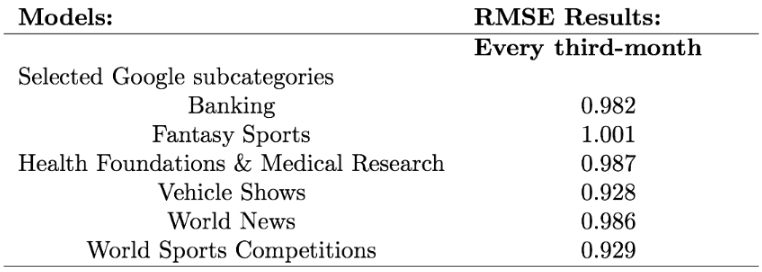
Figure 21: Banking subcategory model and Finland's confidence model

Figure 21 presents that Google searches about Banking decreased after the financial crises. After that, searches have remained stable, which could be because this banking model used three-month average data. Nonetheless, even with the optimal chosen subcategory consumer confidence data seems to be the superior soft data source. The following table 13 presents optimal subcategories for Germany.

Table 13: Germany's subcategory models selected by LASSO shrinkage method



Country:	**Germany**
Shrinkage method:	**LASSO**
Models:	**RMSE Results:**
	Three-months average
Selected Google subcategories	
Aging & Geriatrics	1.062
Banking	0.972
Bus & Rail	1.032
Business Education	1.005
Business News	1.014
Classifieds	1.011
Fantasy Sports	0.979
Gardening & Landscaping	0.975
Hair Care	0.993
Health Conditions	1.026
Local News	0.968
Motor Sports	0.984
Nursery & Playroom	1.000
Public Safety	1.037
Scooters & Mopeds	1.016
Sports Coaching & Training	1.007
Sports News	0.985
Vehicle Codes & Driving Laws	1.013
Vehicle Parts & Accessories	1.049
Vehicle Shows	0.932
Weather	0.978
Weight Loss	0.960
Women's Health	0.994
World News	0.954
World Sports Competitions	0.958



Models:	**RMSE Results:**
	Every third-month
Selected Google subcategories	
Banking	0.982
Fantasy Sports	1.001
Health Foundations & Medical Research	0.987
Vehicle Shows	0.928
World News	0.986
World Sports Competitions	0.929

Table 13 suggests that the driving force behind the Autos & Vehicles category was the Vehicle Shows subcategory. It appears that searches for vehicles shows could be a signal for current GDP. In other words, when people are searching for vehicle shows, they could be planning to purchase a new car. This planning, in turn, could lead to an actual car purchase that would increase Germany's consumption and manufacturing, i.e. mainly the automotive industry. The following figure depicts the optimal Vehicle Shows subcategory model against Germany's GDP growth.

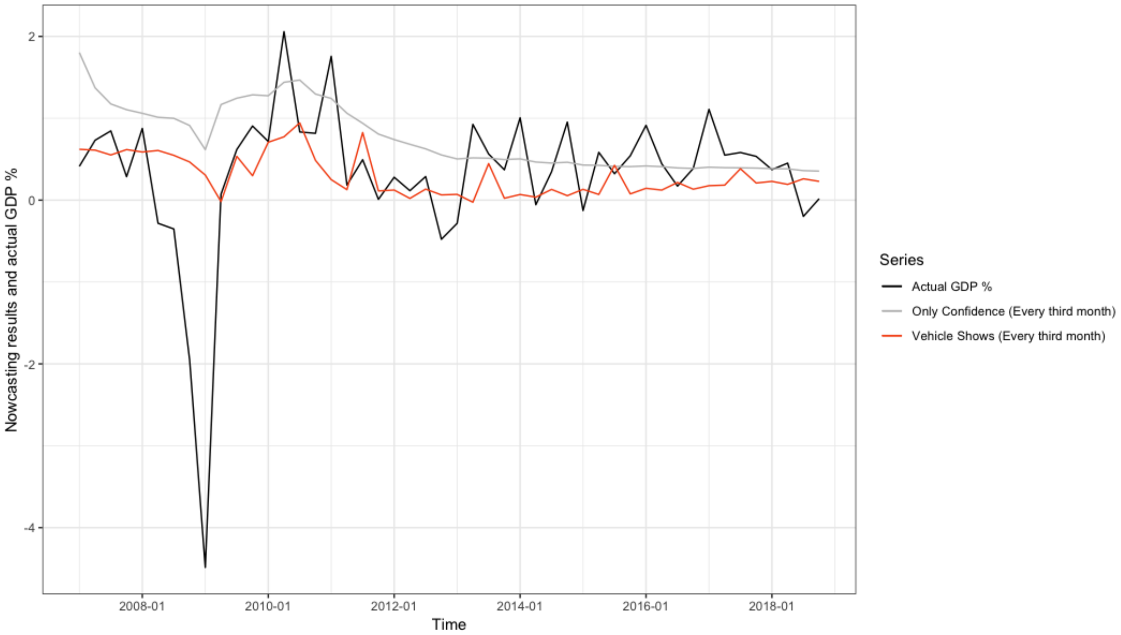


Figure 22: Vehicle Shows subcategory model and Germany's confidence model

Figure 22 implies that Vehicles Shows subcategory model is more in line with Germany's GDP growth than the consumer confidence model. It also suggests that searches for vehicle shows decreased shortly after the financial crises. In addition, there are simultaneous increases in vehicle show searches and GDP. However, careful post-crises examination seems to reveal a cyclical pattern from the subcategory model. Nevertheless, Germany's subcategory models distinctly exceed the consumer confidence model.

Overall, Finland's consumer confidence model outperformed all of the Google subcategory models. Confidence model's results were also more realistic and reliable. In Germany, the opposite was right, as the leading subcategory model unambiguously surpassed the confidence model. This master's thesis also applied model validation techniques to ensure these nowcasting results.

5.1.4 Cross-validation

This thesis applied a *Leave-one-out cross-validation* (LOOCV) method to achieve a more robust analysis of the models' nowcasting performance. Cross-validation methods typically split the data into two subsets the validation data and the training set. However, the leave-one-out cross-validation method follows an iterative process where single observation is repeatedly excluded from the training set. In other words, the leave-one-out cross-validation method is fitting models with training data and creating model predictions with validation data. (James, Tibshirani, Witten & Hastie, 2013, 178–179.)

Unlike previous pseudo-out-of-sample exercises, this thesis's cross-validation models are using "future" data and fixed sample size. This method allows for a longer forecasting period; hence, there are more point forecasts to evaluate and examine. More throughout depiction of this thesis' cross-validation arrangement is in appendix 2.

This thesis applied these methods for both Finland and Germany's leading Google and confidence models. For the leave-one-out cross-validation method, this thesis also included new nowcasting model, which is in equation 22.

$$GDP_t = \beta_0 + \beta_1 Google_{it} + \beta_2 Confidence_t \tag{22}$$

$$t = 1, \dots, \mathrm{T}$$
$$i = 1, \dots, \mathrm{N}$$

Model in equation 22 includes both the country's leading Google category and its consumer confidence data. As previously stated, cross-validation allows for more extended estimation period as the earlier pseudo-of-sample. This extended period should benefit complex models, as shown in equation 22. The following table 14 display Finland's cross-validation results.

Table 14: Finland's leave-one-out cross-validation results

Country:	**Finland**
Model:	**RMSE Result:**
	Every third-month
Only Confidence (16)	1.208

Dimension reduction method:	**Principal component analysis (PCA)**
Models:	**RMSE Results:**
	Three-months average
Only Google categories (17)	
News	1.290
Google categories and Confidence (22)	
News	1.239

As previously stated, Finland's leading Google model was the News category model. Table 14 imply that the cross-validation method still recommends consumer confidence as the most accurate model. Cross-validation method also led to higher differences between the leading Google and confidence models RMSE results.

It is also interesting that Google models forecasting accuracy improved when it included the consumer confidence data. However, with RMSE result of 1.208, the consumer confidence model is still able to prevail. Moreover, Finland's News category models' results did not significantly differ from the intercept-term. Nevertheless, bellow figure 23 depicts Finland's cross-validation estimates for the confidence augmented model and confidence model.

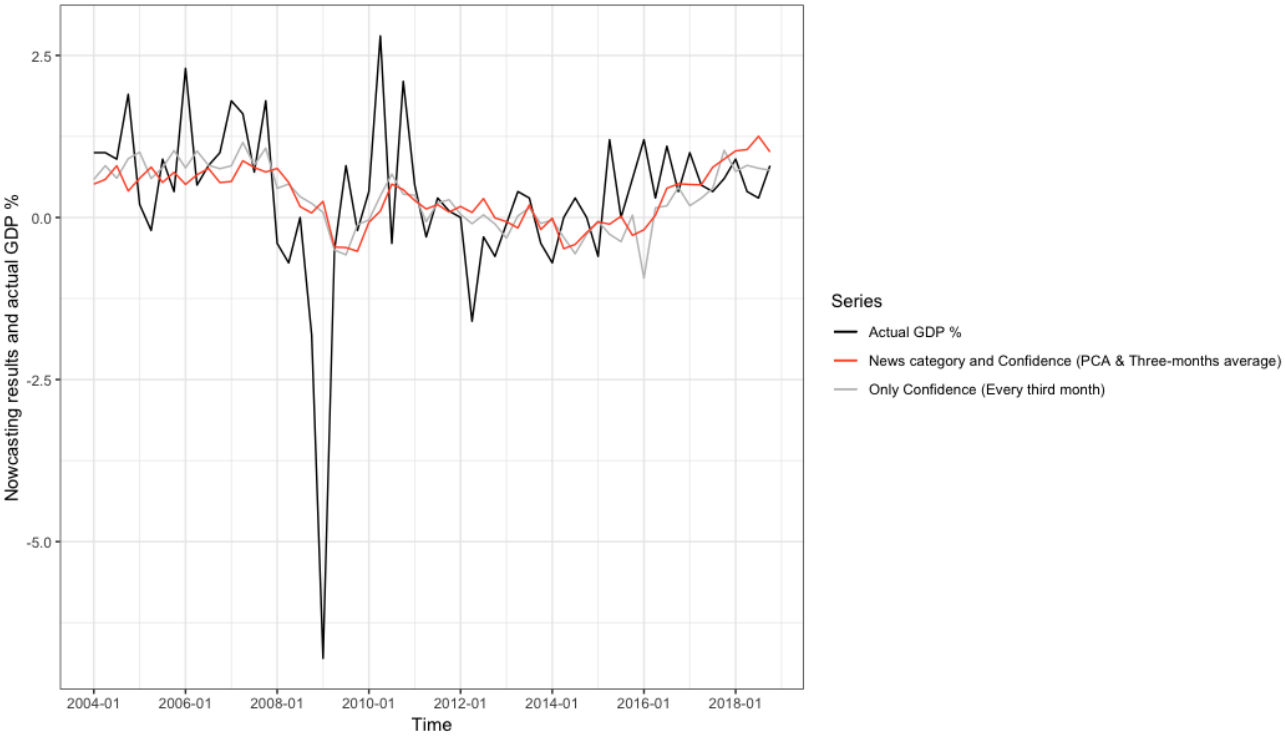


Figure 23: Confidence augmented News category model and Finland's confidence model

As evident from the figure 23, the models are now able to generate more point forecasts. Furthermore, these forecasts are mostly similar. The main difference between them is that the confidence-augmented model's estimates are smoother than the univariate confidence models. This smoothness is because the leading News category uses "three-month average" data. However, this smoothness does not improve the predictive accuracy of the model. Therefore, the univariate confidence model is the most accurate and reliable cross-validation model to nowcast Finland's GDP growth. The following table 15 present RMSE results for Germany's cross-validation models.

Table 15: Germany's leave-one-out cross-validation results

Country:	**Germany**
Model:	**RMSE Result:**
	Every third-month
Only Confidence (16)	0.882

Dimension reduction method:	**Partial least squares (PLS)**
Models:	**RMSE Results:**
	Every third-month
Only Google categories (17)	
News	0.962
Google categories and Confidence (22)	
News	0.893

Compared to Finland, there are smaller differences between the models RMSE result. Table 15 indicate that the cross-validation method provides noticeable gains to confidence models estimation accuracy, i.e. lower RMSE result. Confidence data also improves the Google model's forecasting accuracy as the augmented model had the lowest RMSE result. However, table 15 results imply that univariate Google model had the lowest accuracy.

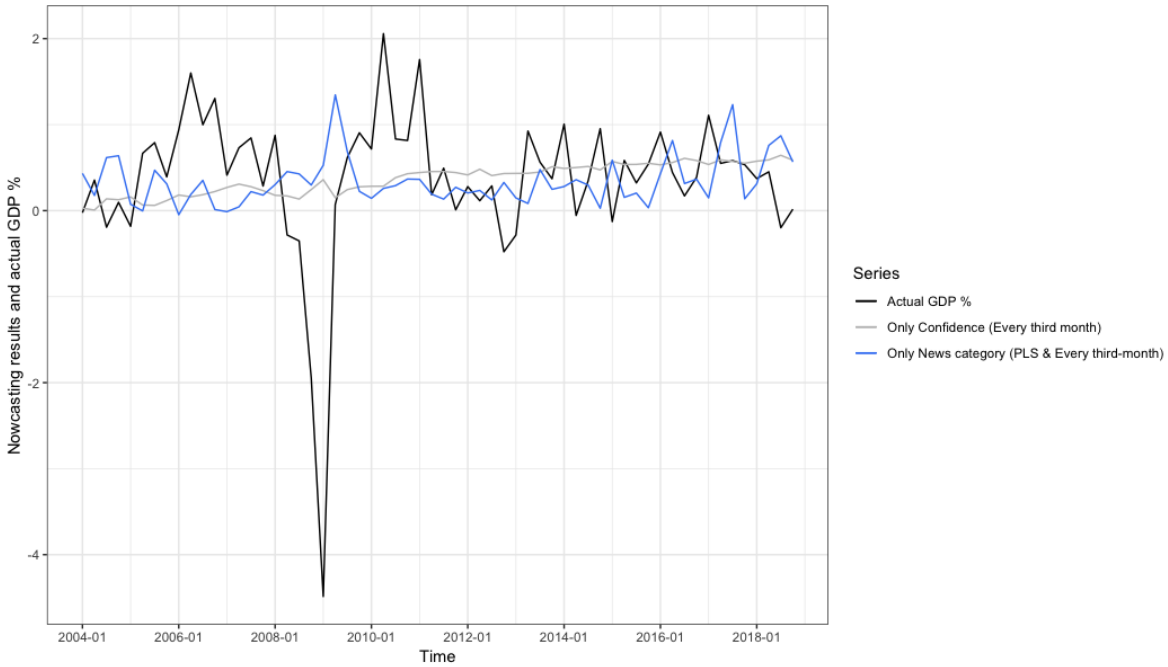


Figure 24: News category model and Germany's confidence model

Figure 24 indicates that Germany's News model is somewhat accurate nowcasting model. Furthermore, the News category model generates considerably more variation than the consumer confidence model. Despite this, the univariate News category model is the most inaccurate cross-validation model in Germany. It is also interesting that despite the different estimation method, the News model still forecasts a significant increase in GDP growth in amidst of the financial crises. This noticeable increase is a potential signal of the Google searches relationship with policy-related uncertainty. Anyhow, following figure 25 present both the confidence and the leading model, i.e. confidence augmented News category model.

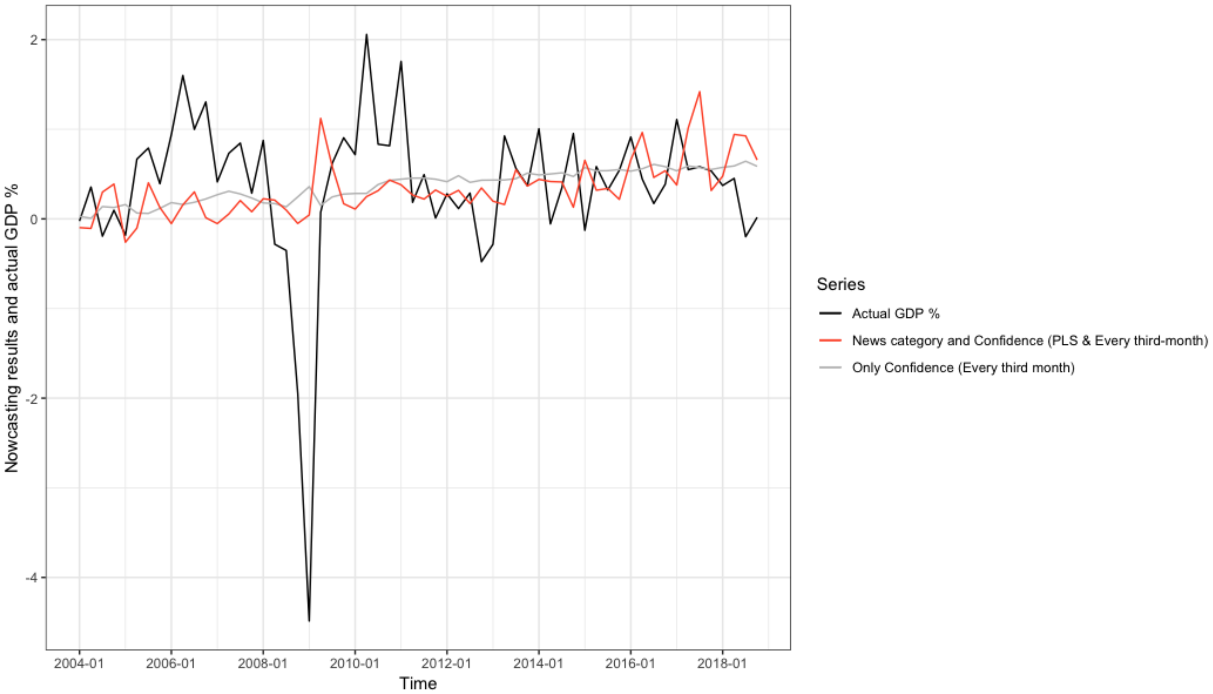


Figure 25: Confidence augmented News category model and Germany's confidence model

Figure 25 presents intriguing estimates produced by the leading cross-validation model, i.e. equation 22. Confidence augmented model cannot only react appropriately to the financial crisis; it also has several adequate responses to Germany's GDP changes. However, the last two years of the estimation period are quite tricky, as the model seem to exaggerate the actual GDP growth. Even so, it appears that a combination of Google and confidence data can create useful nowcasting results.

In conclusion, cross-validation confirms the earlier results that the consumer confidence model was the most accurate model to nowcast Finland's GDP growth. In Germany, cross-validation reinforced confidence augmented Google model's position as the most precise nowcasting model. Thus, Google Trends data works better in Germany than in Finland, where consumer confidence data is a more efficient data source describing the country's GDP growth. However, figures 20 & 24 present some evidence about Google searches relationship with economic uncertainty. Thus, it would be interesting to examine do Google searches have a relationship policy-related uncertainty and how it affects Google models' forecasts.

5.1.5 Google Trends and policy-related uncertainty

Donadelli (2015) found that Google searches have relation to policy-related uncertainty. Furthermore, policy-related Google searches are particularly popular when there are significant levels of uncertainty for economic conditions (Donadelli, 2015, 802). Growing uncertainty also tends to make people more cautious about consuming and investing (Donadelli, 2015, 802). Therefore, when economic conditions are favorable, there should be a considerable number of Google searches for durable goods. This master's thesis includes these uncertain economic conditions in the following equation 23.

(23)
$$GDP_t = \beta_0 + \beta_1 GDP_{it-1} + \beta_2 Google_{it} + \beta_3 Google_{it} * Uncertainty_t + \varepsilon_t$$

$$t = 1, \dots, \mathrm{T}$$
$$i = 1, \dots, \mathrm{N}$$

Equation 23 describes a model, which includes a new interaction term $Google_{it} * Uncertainty_t$. This interaction term's $Uncertainty_t$ represents the United States policy-related uncertainty, which this study has collected from Economic Policy Uncertainty (EPU) database[4]. More specifically, this thesis uses the EPU's news-based policy uncertainty index developed by Baker, Bloom & Davis (2016), which is measured in terms of newspaper coverage frequency. This news-based uncertainty index includes data from the United States 10 largest newspapers, i.e. the index searches newspaper articles containing words regarding uncertainty and politics.

In this manner, the model in equation 23 is controlling for increased news-based policy-related uncertainty in the US and its influence on Finland and Germany's Google searches. It is worth noting that increased policy-related uncertainty might potentially lead to weaker GDP forecasts. This thesis's uncertainty models' results are in the following tables 16 & 17.

[4] Economic Policy Uncertainty (EPU) indexes are available in https://www.policyuncertainty.com

Table 16: RMSE results of Finland's uncertainty model (23)

Country:	**Finland**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 with Google and Uncertainty (23)		
Autos & Vehicles	7.193	4.248
Beauty & Fitness	6.487	3.726
Business & Industrial	8.345	4.954
Computers & Electronics	4.363	3.004
Food & Drink	3.804	2.968
Health	8.491	5.050
Home & Garden	6.895	3.609
Internet	4.739	3.288
Investing	4.243	2.780
Jobs	2.821	2.475
Law	3.853	2.464
News	3.532	3.705
Real Estate	5.466	4.118
Shopping	5.783	3.748
Sports	5.324	3.006
Travel	5.294	3.595

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 with Google and Uncertainty (23)		
Autos & Vehicles	121.087	75.325
Beauty & Fitness	95.953	54.560
Business & Industrial	159.313	83.055
Computers & Electronics	116.101	68.137
Food & Drink	45.652	24.979
Health	156.904	75.156
Home & Garden	117.332	57.897
Internet	83.848	46.652
Investing	59.210	27.989
Jobs	57.994	32.740
Law	77.930	38.485
News	35.956	25.241
Real Estate	71.634	40.148
Shopping	97.778	51.052
Sports	95.591	51.149
Travel	115.704	65.210

Table 16 illustrates US policy-related uncertainty's influence on Finland's Google searches, i.e. the equation 23. Now data averaging has a noticeable disadvantage, which is because uncertainty is a more volatile indicator. Additionally, table 14 suggests that the PLS method is a hindrance when models include the uncertainty term. This issue may be because PLS reduces Google Trends data's dimension based on the covariance with GDP. Mainly, it produces information sets that have high covariance with Finland's GDP. Thus, it seems that Finland's GDP growth is not that related to US news-based policy uncertainty.

However, PCA methods results imply that Finland's Google searches are associated with US policy-related uncertainty. This possible is because the PCA method does not regard Finland's GDP when transforming the information sets. Thus, it appears that Google Trends categories have some tendency to respond to macroeconomic policy-related uncertainties.

Nonetheless, according to the RMSE results, the most accurate models were the Investing, Jobs and Law category models. It seems that when people feel uncertain about the economic conditions, the searches regarding finance, jobs and legislations are greatly affected. This result is somewhat intuitive as these categories of information should be in people's interests in high levels of uncertainty. The uncertainty models nowcasting estimates are in the subsequent figures 27 and 28.

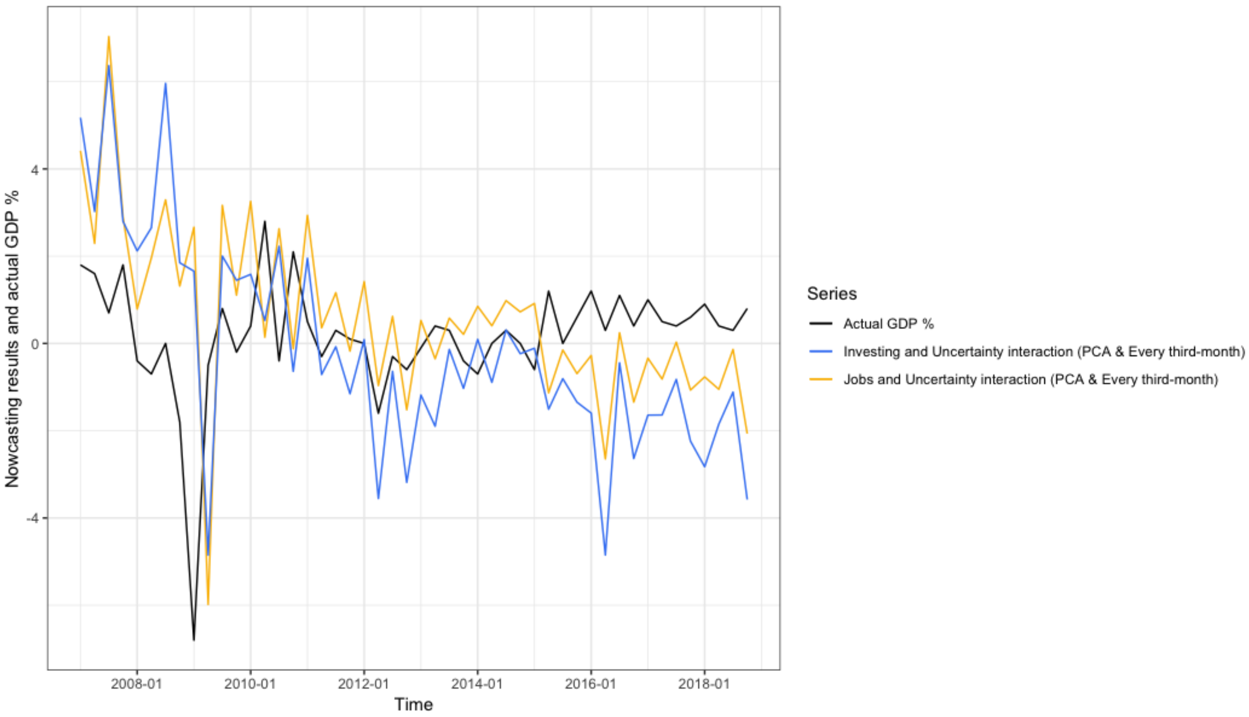


Figure 26: Job, Investing and Uncertainty models against Finland's GDP growth

Difference to previous figures, figure 26 shows highly volatile nowcasting estimates. Higher volatility is due to the uncertainty data, which is quite volatile. Nonetheless, Jobs and Investing categories seem to move in unison.

Most noticeable movements being the surges in 2007 and 2009. Although there are some confluences with Finland's GDP, the uncertainty models with Jobs and Investment do not produce reliable nowcasting results.

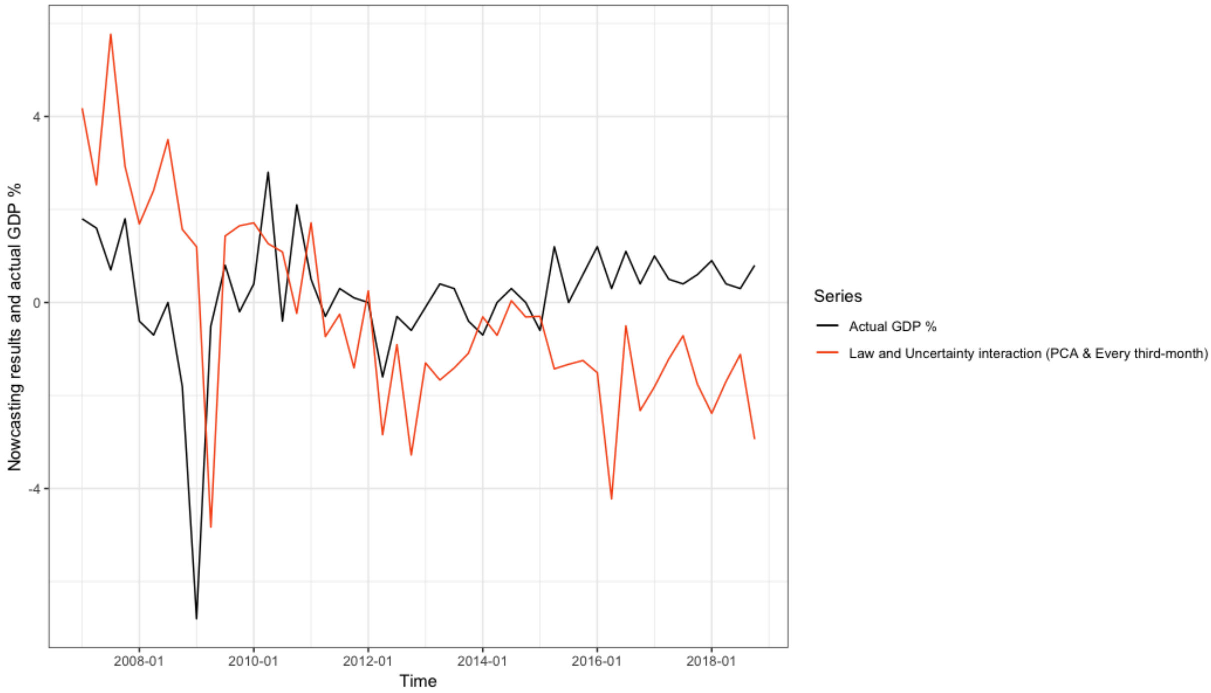


Figure 27: Food & Drink and Uncertainty models against Finland's GDP growth

The leading uncertainty model included the Law Google Trends category. When looking at figure 27, it appears that the model nowcasts a considerably lower GDP growth than the actual GDP. Furthermore, it seems to forecast a noticeable decrease to Finland's GDP in late 2018.

Regardless, provided by the results of table 16, figures 26 and 27, it appears that Google Trends categories do not work well with policy-related uncertainty data, at least for Finland's GDP. News based uncertainty data generates too much volatility to the nowcasting models. Higher volatility, in turn, leads to the models to aggravate the GDP changes in a way that is not plausible in real-life. The following table 17 presents Germany's uncertainty models RMSE results.

Table 17: RMSE results of Germany's uncertainty model (23)

Country:	**Germany**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 with Google and Uncertainty (23)		
Autos & Vehicles	4.363	2.369
Beauty & Fitness	3.122	2.202
Business & Industrial	5.555	2.764
Computers & Electronics	3.302	2.201
Food & Drink	2.822	3.017
Health	5.591	3.030
Home & Garden	4.205	2.309
Internet	2.977	2.126
Investing	2.558	1.871
Jobs	1.564	1.243
Law	2.540	1.706
News	3.285	1.666
Real Estate	2.732	1.654
Shopping	3.897	3.437
Sports	3.421	1.625
Travel	2.319	1.736

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 with Google and Uncertainty (23)		
Autos & Vehicles	17.346	8.077
Beauty & Fitness	21.810	19.773
Business & Industrial	62.709	30.698
Computers & Electronics	72.718	40.174
Food & Drink	13.338	16.154
Health	39.365	15.245
Home & Garden	14.565	9.788
Internet	32.631	20.242
Investing	32.863	21.971
Jobs	27.231	14.025
Law	38.979	16.943
News	13.069	3.895
Real Estate	28.064	14.848
Shopping	44.830	37.700
Sports	20.107	8.103
Travel	31.878	12.166

Table 17 describes the US policy-related uncertainty's influence on Germany's Google searches. The table's results look similar to the previous Finnish uncertainty models. As before, the PLS method is producing information sets that have high covariance with Germany's GDP. Therefore, it seems that Germany's GDP does not have a relationship with US news-based policy uncertainty. Also, like Finland's results, PCA methods suggest that also Germany's Google searches relate with US policy-related uncertainty.

As reported in table 17, the most accurate German uncertainty models were Real Estate, Sports and Jobs. In other words, policy-related uncertainty appears to affect people's Google searches for jobs, housing and athletics. The Job category is identical to previous Finnish uncertainty results. However, in Germany, real estate and sports searches are also affected. These two categories are in the following figure 28.

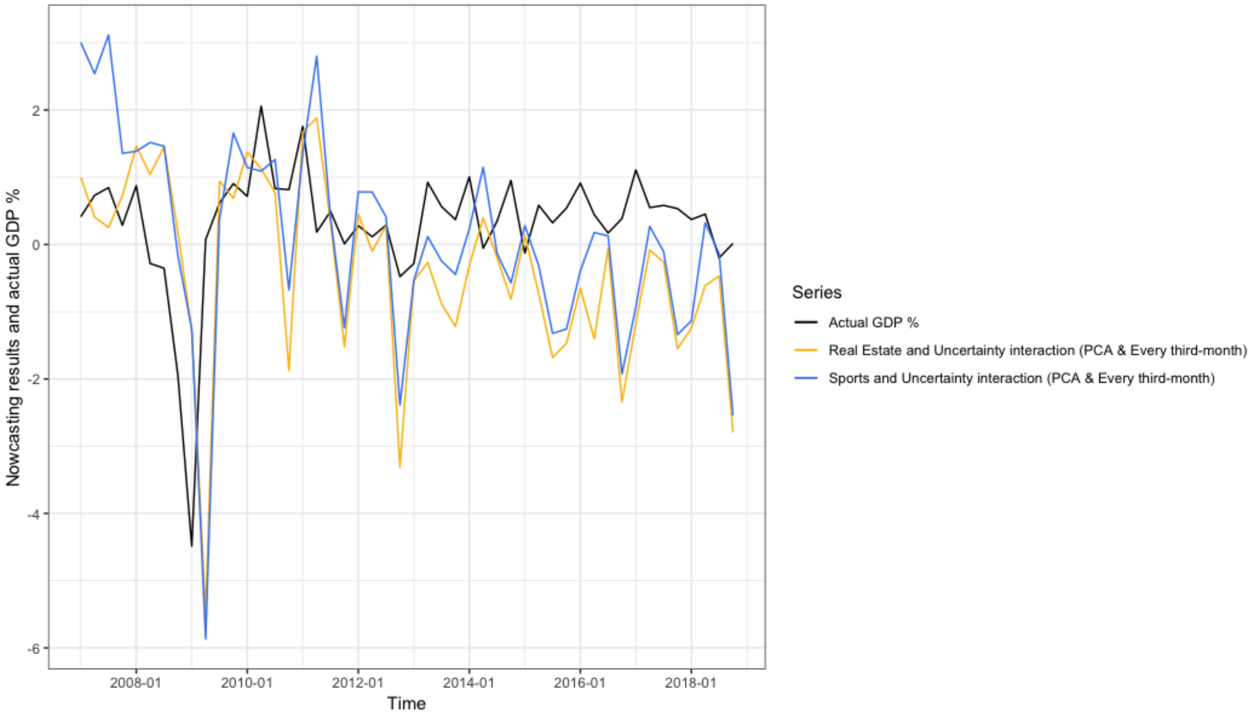


Figure 28: Sports, Travel and Uncertainty models against Germany's GDP growth

As shown in figure 28, the Real Estate category's nowcasts are highly volatile. Although the Sports category was second-most accurate model, it also produced too volatile nowcasts. Moreover, two of the leading uncertainty models do not seem to coincide to Germany's GDP growth. Bellow figure 29 depicts Germany's most accurate uncertainty model.

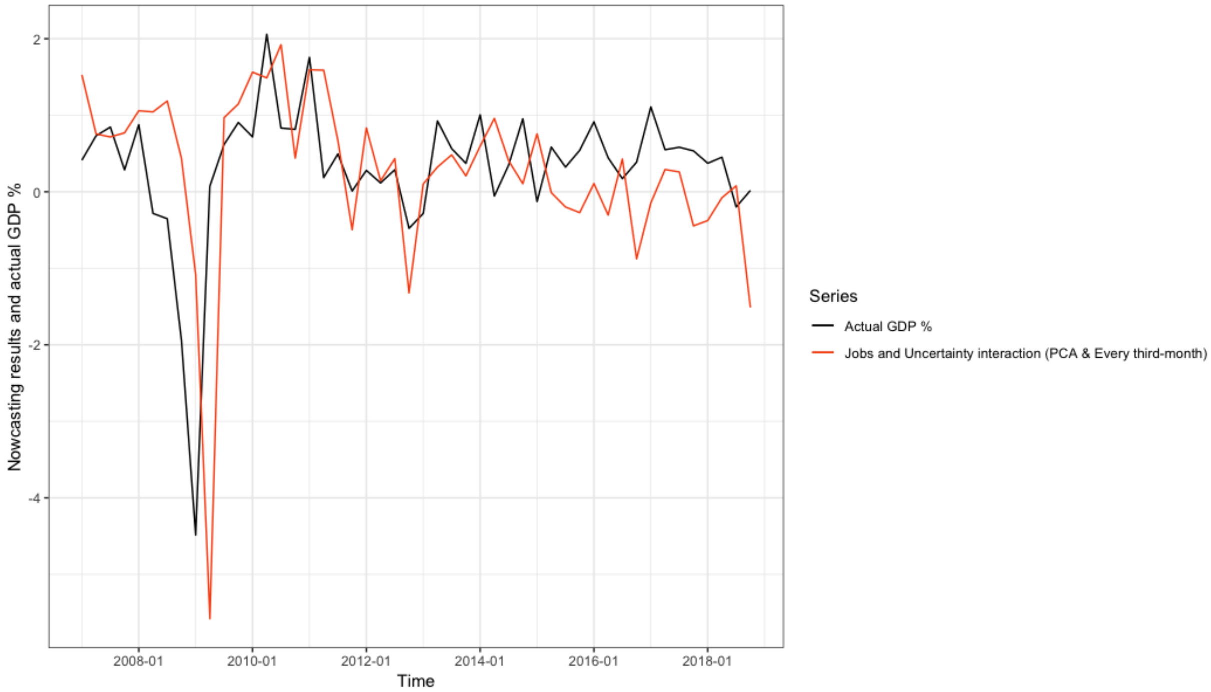


Figure 29: Job and Uncertainty models against Germany's GDP growth

The most accurate Germany's uncertainty model included Google's Job category. Surprisingly enough, Jobs uncertainty model seem to be also this thesis' the most accurate uncertainty model. Its nowcasting results mostly coincide with Germany's GDP growth. Despite these positive results, the Jobs category model's estimates are also too volatile to use in practice.

Overall, the uncertainty models are not effectively nowcasting either of the country's GDP growth. However, both countries' results suggest that people's searches for Jobs are one of the most affected searches when US policy-related uncertainty arises. In Finland, law and investing related searches are also affected. In Germany, US policy-related uncertainty affected searches regarding people's real estate planning and sports.

5.2 Discussion of the results

According to the results of this master's thesis, the leading nowcasting models used only one variable. Results also suggest that Google Trends data models were generally able to outperform the benchmark (AR-1) models. Broad category models also confirmed this. In Finland, the most accurate Google Trends category models were constructed using principal component analysis (PCA) dimension reduction method and "three-month average" data. Three of the leading broad categories were News, Real Estate and Jobs. However, the consumer confidence model produced the lowest RMSE score.

On the contrary to Finland, most of Germany's leading broad categories were constructed using partial least squares (PLS) and "every third-month" data. These leading categories included News, Autos & Vehicles, Real Estate and Sports. In addition, all these categories were able to surpass the consumer confidence model.

When studying the Google Trends subcategories, LASSO found several optimal subcategories relating to Finland's GDP growth, the most accurate being the Banking category. Even with this selection method, the consumer confidence model proved to be the most accurate nowcasting model. In Germany, LASSO revealed the Vehicle Shows subcategory to be the driving force behind the leading Autos & Vehicles broad category.

This thesis applied leave-one-out cross-validation (LOOCV) to achieve a more comprehensive analysis of Google searches nowcasting abilities. Finland's results further strengthened consumer confidence models role as the leading nowcasting model. What is more, Google models forecasting accuracy improved when it included the confidence data. Germany's leading Google model was constructed using the PLS method, which could potentially be influenced by future GDP. This thesis constructed cross-validation method differently; hence, it provides further evidence about the true nature of the Google data.

Cross-validation results reveal that also in Germany; the consumer confidence can produce the lowest RMSE scores, i.e. it is the most accurate model in the analysis. However, when looking at the figures, confidence models' estimates are far from being as adequate, say, as in Finland. It is also interesting that Google model accuracy once again increases when it includes consumer confidence data.

Donadelli (2015) found that Google search data had a relationship with policy-related uncertainty. This master's thesis used US policy-related uncertainty index to examine if Google searches relate to uncertain economic conditions. Finland's uncertainty models result imply that leading Google categories were Investing, Jobs and Law. Germany's leading uncertainty models included Real Estate, Sport and Jobs. The most accurate being the Jobs category, which generated somewhat surprising results. Nonetheless, all of these models produced highly volatile nowcasting estimates.

5.2.1 Comparing results to earlier studies

Earlier studies, most notably, Götz and Knetsch (2019), Ferrara and Simoni (2019) found that Google Trends data are able to generate additional information about GDP growth. Götz and Knetsch (2019, 21) concluded that in the absence of any official information, Google Trends data could augment models with useful information regarding the German GDP. Ferrara and Simoni (2019) found Google Trends data models to provide accurate forecasts in the first four weeks.

This master's thesis cautiously confirms some of these results as most of the Google models could outperform the benchmark (i.e. AR-1) model. In Finland, consumer confidence models consistently outperform Google Trends models. In other words, there is not a single Google model that had a lower RMSE score than consumer confidence. Thus, in the absence of official data, i.e. Google Trends data manages to produce additional information about Finland's GDP growth.

However, when carefully studying the Finnish nowcasting estimates through figures, consumer confidence models show their dominance. Consumer confidence model's nowcasts, not only foreshadow 2008 crises, but they also moved unison with Finland's actual GDP. This movement is not evident in the leading Google Trends categories, which estimates do not seem to have co-movement with actual GDP. Furthermore, Google Trends categories models' estimates appeared to be much smoother than consumer confidence. Therefore, although Google Trends data can exceed the benchmark model, consumer confidence model produces more robust nowcasting results. This result partially confirms Bańbura & Rünstler (2011) and Giannone et al. (2008) findings of survey data's usefulness in nowcasting.

In Germany, almost all the univariate Google models are able to surpass the confidence model. Moreover, the five leading nowcasting models included only Google category models. The figures also strengthen the Google models supremacy as they produced estimates that coincided with Germany's GDP. On the other hand, Germany's consumer confidence model generated too smooth nowcasting estimates. One of the reasons for this is that consumer confidence data do not relate Germany's GDP as strongly as in Finland. Google models can also generate more accurate results in Germany for another reason. It may be that the Google LLC category algorithm works better in larger countries than in smaller countries, e.g. Finland.

This master's thesis is unable to find conclusive results about Google Trends data's relationship with policy-related uncertainty. Comparison between PLS's and PCA's RMSE results reveal that Google Trends data is possibly associated with policy-related uncertainty. However, the uncertainty model's estimates appear to be highly volatile. This volatility is not plausible in real-life GDP cycles. Furthermore, these estimates are not moving at unison with Finland's GDP growth. Thus, this study is unable to find Finnish Google searches relating to policy-related uncertainty. In Germany, this thesis found some relationship with policy-related uncertainty and Google Trends data. Nevertheless, that relationship requires further examination.

5.2.2 Reliability of the results

Results of this master's thesis suggest that consumer confidence data is superior for nowcasting Finland's GDP. Finland is a relatively small country; therefore; its GDP is highly dependent on exports. Thus, Finnish Google searches could have difficulties capturing relevant information about the economy. Consequently, Google Trends data seem to work better in Germany, which is a large country with a large manufacturing sector. However, most of Germany's leading models were constructed using the partial least squares (PLS) methods, in which ex-post GDP potentially influencing the forecasts. Moreover, cross-validation results suggest that although Germany's consumer confidence model was not able to follow the GDP that carefully, it had the lowest RMSE score.

Additionally, Google Trends data were relatively short beginning in January 2004, which significantly limits the possible estimation length. Still, the quantity of searches has increased considerably, and it is safe to assume that the search terms have also changed. In other words, Google searches data may have changed substantially throughout the years. Therefore, in the future, as Google Trends data grows, and depending on how Google's algorithm evolves, it might have a better representation of people's interests and preferences.

This master's thesis Google Trends data were somewhat similar to Götz and Knetsch (2019). This thesis included the same initial broad and subcategories. However, Götz and Knetsch (2019) used private ECB Google Trends data, which is formed differently than the publicly available data. Despite this, the overall results did not differ significantly.

This thesis's methods followed Götz and Knetsch (2019) study as it applied similar dimension reduction methods, i.e. principal component analysis and partial least squares. This thesis also used partially same LASSO shrinkage method as in Götz and Knetsch (2019). However, Götz and Knetsch (2019) applied bridge equation models. This master's thesis examined more straightforward nowcasting models. Furthermore, it is worth noting that there are minor pitfalls in these methods. For example, LASSO may produce unstable estimates (Lim & Yu, 2016). Nevertheless, these research methods are relevant in terms of answer the research question.

In this context, one may also consider if Google Trends data is relevant for nowcasting GDP growth. It may be better suited to nowcast different consumer confidence categories, e.g. housing, cars and finance. Likewise, consumer confidence is more closely associated with people's interests than countries GDP growth. Even so, this study found that Google Trends data has a minor correlation with consumer confidence. That evidence suggests that they both have some similar information within them, but this related information was not that evident in the nowcasting estimates.

6 CONCLUSIONS

This master's thesis attempted to nowcast Finland and Germany's GDP growth using Google Trends. To answer the question of whether Google Trends data is any good. The results suggest that Google data is, in fact, able to generate additional information in both countries as it outperforms the benchmark model. Still, careful examination reveals that Finland's consumer confidence models are consistently superior to Google models.

In Germany, the five leading Google models surpassed the consumer confidence model. However, cross-validation analysis revealed that consumer confidence model could produce forecasts that are more accurate than the leading Google model. This master's thesis also investigated Google Trends data's relationship with policy-related uncertainty, but there was no conclusive evidence supporting this argument. Differences in dimension reduction results confirm relationship with uncertainty, but nowcasting estimates were too volatile.

Nevertheless, there are still multiple possibilities for further studies. Further studies could study Google Trends data relationship with policy-related uncertainty in other large developed countries, e.g. the United States, the United Kingdom. Furthermore, low-income developing countries could present an intriguing research topic as their official GDP statistics are difficult to produce.

In addition, as previously stated, Google Trends data is not in the absolute numeric form, and this thesis applied only recursive nowcasting exercise. Therefore, further studies could use the Kalman filter to estimate dynamic nowcasting models with the actual values of the search terms. Moreover, future models using Google Trends could focus on nowcasting turning points in GDP growth.

REFERENCES

Anttonen, J. (2018). Nowcasting the unemployment rate in the EU with seasonal BVAR and google search data. *(No. 62). ETLA Working Papers.*

Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly, 55*(2), 107-120.

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

Bańbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-casting and the real-time data flow. *Handbook of Economic Forecasting, 2*(Part A), 195-237.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Nowcasting. *ECB Working Paper No. 1275, European Central Bank*.

Bańbura, M., & Rünstler, G. (2011). A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting, 27*(2), 333-346.

Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial Markets and Portfolio Management, 25*(3), 239.

Choi, H., & Varian, H. (2009a). Predicting initial claims for unemployment benefits. *Technical Report, Google Inc*, 1-5.

Choi, H., & Varian, H. (2009b). Predicting the present with google trends. *Technical Report, Google Inc*, 1-5.

Choi, H., & Varian, H. (2012). Predicting the present with google trends. *Economic Record, 88*, 2-9.

Cooper, C. P., Mallon, K. P., Leadbetter, S., Pollack, L. A., & Peipins, L. A. (2005). Cancer internet search activity on a major search engine, united states 2001-2003. *Journal of Medical Internet Research, 7*(3).

D'Amuri, F., & Marcucci, J. (2009). 'Google it!' forecasting the US unemployment rate with a google job search index. *ISER Working Paper Series, 2009-32*.

Della Penna, N., & Huang, H. (2009). Constructing consumer sentiment index for US using google searches. *Working Papers No. 2009-26, Edmonton: University of Alberta, Department of Economics.*

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

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

European Commission. (2019). Consumer confidence indicator. Retrieved from https://ec.europa.eu/info/business-economy-euro/indicators-statistics/economic-databases/business-and-consumer-surveys/download-business-and-consumer-survey-data/time-series_en#consumers

Eurostat. (2018). Internet access and use statistics. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Internet_access_and_use_statistics_-_households_and_individuals&oldid=379591

Evans, M. (2005). Where are we now? Real-time estimates of the macro economy. *NBER Working Paper Series, w11064.*

Ferrara, L., & Simoni, A. (2019). When are google data useful to nowcast GDP? an approach via pre-selection and shrinkage. *Banque de France Working paper April 2019, WP #717.*

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

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature, 457*(7232), 1012.

Google. (2008). Announcing google insights for search. Retrieved from https://adwords.googleblog.com/2008/08/announcing-google-insights-for-search.html

Google. (2010). Google instant, behind the scenes. Retrieved from https://googleblog.blogspot.com/2010/09/google-instant-behind-scenes.html

Google. (2019a). Explore results by region. Retrieved from https://support.google.com/trends/answer/4355212?hl=en&ref_topic=4365530

Google. (2019b). Google trends support. Retrieved from https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052

Google. (2019c). Refine Trends results by category. Retrieved from https://support.google.com/trends/answer/4359597?hl=en&ref_topic=4365530

Götz, T. B., & Knetsch, T. A. (2019). Google data in bridge equation models for german GDP. *International Journal of Forecasting, 35*(1), 45-66.

Hamilton, J. D. (2011). Calling recessions in real time. *International Journal of Forecasting, 27*(4), 1006-1026.

Itkonen, J., & Juvonen, P. (2017). Nowcasting the Finnish economy with a large bayesian vector autoregressive model. *BoF Economics Review 6/2017.*

James, G., Tibshirani, R., Witten, D., & Hastie, T. (2013). *An introduction to statistical learning: With applications in R*. New York: Springer.

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed ed.). New York: Springer. Retrieved from https://jyu.finna.fi/Record/jykdok.968252

Kholodilin, K. A., Podstawski, M., & Siliverstovs, B. (2010). Do google searches help in nowcasting private consumption? A real-time evidence for the US. *KOF Working Papers, 256.*

Koop, G., & Onorante, L. (2013). Macroeconomic nowcasting using google probabilities. *Working Paper, University of Strathclyde.*

Lim, C., & Yu, B. (2016). Estimation stability with cross-validation (ESCV). *Journal of Computational and Graphical Statistics, 25*(2), 464-492.

Martínez-Galán, E., & Artola, C. (2012). Tracking the future on the web: Construction of leading indicators using internet searches. *Banco De Espana Occasional Paper, (1203).*

McLaren, N., & Shanbhogue, R. (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin, 51*(2), 134-140.

OECD. (2019). Quarterly GDP statistics. Retrieved from https://data.oecd.org/gdp/quarterly-gdp.htm#indicator-chart

Pentland, A. (2010). *Honest signals: How they shape our world*. Cambridge, MA: MIT Press.

Perlin, M. S., Caldeira, J. F., Santos, A. A. P., & Pontuschka, M. (2017). Can we predict the financial markets based on Google's search queries? *Journal of Forecasting, 36*(4), 454-467.

Preis, T., Reith, D., & Stanley, H. E. (2010). Complex dynamics of our economic life on different scales: Insights from search engine query data. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 368*(1933), 5707-5719.

Reichlin, L., Giannone, D., & Doz, C. (2006). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *No. 2006-23, THEMA Working Papers.*

Rosipal, R., & Krämer, N. (2005). (2005). Overview and recent advances in partial least squares. Paper presented at the *International Statistical and Optimization Perspectives Workshop" Subspace, Latent Structure and Feature Selection"*, 34-51.

Rünstler, G., & Sedillot, F. (2003). Short-term estimates of euro area real GDP by means of monthly data. *(No. 276), ECB Working Paper.*

Statista. (2018). Market share of Google. Retrieved from https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/

Statistics Finland. (2007). Quality guidelines for official statistics. Retrieved from http://tilastokeskus.fi/org/periaatteet/laatuatilastoissa_en.html

Statistics Finland. (2016). Finnish residents use the internet more and more often. Retrieved from https://www.stat.fi/til/sutivi/2016/sutivi_2016_2016-12-09_tie_001_en.html

Statistics Finland. (2017). Consumer survey: Methodological description. Retrieved from https://www.stat.fi/til/kbar/kbar_2017-05-05_men_001_en.html

Statistics Finland. (2019a). Consumer confidence March 2019. Retrieved from https://www.stat.fi/til/kbar/2019/03/kbar_2019_03_2019-03-27_tie_001_en.html

Statistics Finland. (2019b). Quarterly national accounts 2018. Retrieved from http://www.stat.fi/til/ntp/2018/04/ntp_2018_04_2019-02-28_tie_001_fi.html

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

Stock, J. H., & Watson, M. W. (2008). Phillips curve inflation forecasts. *(No. w14322). National Bureau of Economic Research.*

Suhoy, T. (2009). Query indices and a 2008 downturn: Israeli data. *Bank of Israel Working Papers 2009.06, Bank of Israel.*

Trehan, B. (1989). Forecasting growth in current quarter real GNP. *Economic Review-Federal Reserve Bank of San Francisco,* (1), 39.

Tuhkuri, J. (2014). Big data: Google searches predict unemployment in Finland. *ETLA Reports, 31*.

Veldhuizen, S. v., Vogt, B., & Voogt, B. (2016). Internet searches and transactions on the Dutch housing market. *Applied Economics Letters, 23*(18), 1321-1324.

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

Vosen, S., & Schmidt, T. (2012). A monthly consumption indicator for Germany based on internet search query data. *Applied Economics Letters, 19*(7), 683-687.

World Bank. (2019a). Individuals using the internet (% of population). Retrieved from https://data.worldbank.org/indicator/IT.NET.USER.ZS?end=2017&start=2004&view=chart

World Bank. (2019b). Mobile cellular subscriptions. Retrieved from https://data.worldbank.org/indicator/IT.CEL.SETS?end=2017&start=2004&view=chart

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

APPENDIX 1 Initial Google Trends subcategories

Table 1.1: Initial subcategories 1

Broad categories	Subcategories	Broad categories	Subcategories
Autos & Vehicles		Beauty & Fitness	
	Bicycles & Accessories		*Beauty Pageants*
	Boats & Watercraft		*Body Art*
	Campers & RVs		*Cosmetology & Beauty Professionals*
	Classic Vehicles		*Cosmetic Procedures*
	Commercial Vehicles		*Face & Body Care*
	Custom & Performance Vehicles		*Fashion & Style*
	Hybrid & Alternative Vehicles		*Fitness*
	Microcars & City Cars		*Hair Care*
	Motorcycles		*Spas & Beauty Services*
	Off-Road Vehicles		*Weight Loss*
	Personal Aircraft		
	Scooters & Mopeds	Computers & Electronics	
	Trucks & SUVs		*CAD & CAM*
	Vehicle Brands		*Computer Hardware*
	Vehicle Codes & Driving Laws		*Computer Security*
	Vehicle Maintenance		*Consumer Electronics*
	Vehicle Parts & Accessories		*Electronics & Electrical*
	Vehicle Shopping		*Enterprise Technology*
	Vehicle Shows		*Networking*
			Programming
Business & Industrial			*Software*
	Advertising & Marketing		
	Aerospace & Defense	Investing	
	Agriculture & Forestry		*Accounting & Auditing*
	Automotive Industry		*Banking*
	Business Education		*Credit & Lending*
	Business Finance		*Financial Planning*
	Business Operations		*Grants & Financial Assistance*
	Business Services		*Insurance*
	Chemicals Industry		*Investing*
	Construction & Maintenance		
	Energy & Utilities	Food & Drink	
	Hospitality Industry		*Alcoholic Beverages*
	Industrial Materials & Equipment		*Cooking & Recipes*
	Manufacturing		*Grocery & Food Retailers*
	Metals & Mining		*Non-Alcoholic Beverages*
	Pharmaceuticals & Biotech		*Restaurants*
	Printing & Publishing		
	Professional & Trade Associations	Health	
	Retail Trade		*Aging & Geriatrics*
	Small Business		*Alternative & Natural Medicine*
	Textiles & Nonwovens		*Health Conditions*
	Transportation & Logistics		*Health Education & Medical Training*
			Health Foundations & Medical Research
Home & Garden			*Medical Devices & Equipment*
	Bed & Bath		*Medical Facilities & Services*
	Domestic Services		*Medical Literature & Resources*
	Gardening & Landscaping		*Men's Health*
	Home Appliances		*Mental Health*
	Home Furnishings		*Nursing*
	Home Improvement		*Nutrition*
	Home Storage & Shelving		*Oral & Dental Care*
	Homemaking & Interior Decor		*Pediatrics*
	HVAC & Climate Control		*Pharmacy*
	Kitchen & Dining		*Public Health*
	Laundry		*Reproductive Health*
	Nursery & Playroom		*Substance Abuse*
	Pest Control		*Vision Care*
	Swimming Pools & Spas		*Women's Health*
	Yard & Patio		

Table 1.2: Initial subcategories 2

Broad categories	**Subcategories**	**Broad categories**	**Subcategories**
Internet & Telecom		Jobs & Education	
	Communications Equipment		*Education*
	Email & Messaging		*Jobs*
	Mobile & Wireless		
	Search Engines	News	
	Service Providers		*Broadcast & Network News*
	Teleconferencing		*Business News*
	Web Apps & Online Tools		*Gossip & Tabloid News*
	Web Portals		*Health News*
	Web Services		*Journalism & News Industry*
			Local News
Law & Government			*Newspapers*
	Government		*Politics*
	Legal		*Sports News*
	Military		*Technology News*
	Public Safety		*Weather*
	Social Services		*World News*
Shopping		Real Estate	
	Antiques & Collectibles		*Apartments & Residential Rentals*
	Apparel		*Commercial & Investment Real Estate*
	Auctions		*Property Development*
	Classifieds		*Property Inspections & Appraisals*
	Consumer Resources		*Property Management*
	Entertainment Media		*Real Estate Agencies*
	Gifts & Special Event Items		*Real Estate Listings*
	Luxury Goods		*Timeshares & Vacation Properties*
	Mass Merchants & Department Stores		
	Photo & Video Services	Sports	
	Shopping Portals & Search Engines		*College Sports*
	Swap Meets & Outdoor Markets		*Combat Sports*
	Tobacco Products		*Extreme Sports*
	Toys		*Fantasy Sports*
	Wholesalers & Liquidators		*Individual Sports*
			Motor Sports
Travel			*Sporting Goods*
	Air Travel		*Sports Coaching & Training*
	Bus & Rail		*Team Sports*
	Car Rental & Taxi Services		*Water Sports*
	Carpooling & Ridesharing		*Winter Sports*
	Cruises & Charters		*World Sports Competitions*
	Hotels & Accommodations		
	Luggage & Travel Accessories		
	Specialty Travel		
	Tourist Destinations		
	Travel Agencies & Services		
	Travel Guides & Travelogues		

APPENDIX 2 Gradual description of the nowcasting exercises

This master's thesis **pseudo-out-of-sample** forecasting exercises were conducted in the following order:

1) Data preparation
    a. Monthly data series were adapted to quarterly data
        i. Three-month averaging
        ii. Every third-month data
    b. Relevant variables were reconciled in one data matrix
        i. Data covers the years from 2004 to 2018
        ii. Data's length was 60 periods
2) Loop procedure
    a. Initial sample size length was 12 periods
    b. Loop ran through the entire data so that the model's sample was continually increasing, and each observation was used in the estimation.
    c. If the model included Google Trends data, dimension reduction method was conducted, i.e. principal component analysis (PCA) or partial least squares (PLS). Selected dimension reduction method produced a single Google variable, which was implemented to the model
    d. Each model in question was estimated with ordinary least squares (OLS)
3) Nowcasts were created by multiplying the "original data" with model estimates, i.e. the model's coefficients
    a. PCA and PLS methods "original data" were created with ex-post Google data
        i. However, in PLS method the "original data" for the Google variable were constructed with the respect of the ex-post GDP
4) Root mean squared errors (RMSE) results were generated comparing the model forecasts and the actual GDP data

Thesis's **leave-one-out cross-validation** exercises had the following structure:

1) Data preparation
    a. Monthly data series were adapted to quarterly data
        i. Three-month averaging
        ii. Every third month
    b. Relevant variables were reconciled in one data matrix
        i. Data covers the years from 2004 to 2018
        ii. Data's length was 60 periods
2) Loop procedure
    a. Cross-validation used all data, but each loop iteration removed one row of data. In other words, models could not use the country's current GDP in their estimates.
    b. If the model included Google Trends data, dimension reduction method was conducted, i.e. principal component analysis (PCA) or partial least squares (PLS). Selected dimension reduction method produced a single Google variable, which was implemented to the model.
    c. Each model in question was estimated with ordinary least squares (OLS)
3) Forecasts were created by multiplying the "original data" with model estimates, i.e. the model's coefficients
    a. In cross-validation, the "original data" is the first column of the score-matrix, which is formed from a simple matrix calculation
        i. It is done by first taking the inverse of the cross-validation's PLS loadings matrix and by multiplying it with the Google data. Second, this result is then subtracted by its mean.
4) Root mean squared errors (RMSE) results were generated comparing the model forecasts and the actual GDP data

APPENDIX 3 Multivariate models' results

Table 2.1: RMSE results of Finland's model (19)

Country:	**Finland**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 and Google categories (19)		
Autos & Vehicles	1.605	1.656
Beauty & Fitness	1.664	1.729
Business & Industrial	1.645	1.654
Computers & Electronics	1.658	1.690
Food & Drink	1.692	1.764
Health	1.672	1.695
Home & Garden	1.586	1.641
Internet	1.688	1.708
Investing	1.689	1.714
Jobs	1.670	1.716
Law	1.634	1.691
News	1.685	1.766
Real Estate	1.599	1.682
Shopping	1.650	1.698
Sports	1.639	1.643
Travel	1.658	1.684

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 and Google categories (19)		
Autos & Vehicles	1.604	1.690
Beauty & Fitness	1.805	1.805
Business & Industrial	1.870	1.742
Computers & Electronics	1.673	1.694
Food & Drink	1.698	1.724
Health	1.795	1.801
Home & Garden	1.637	1.680
Internet	1.694	1.723
Investing	1.674	1.700
Jobs	1.846	1.805
Law	1.661	1.695
News	1.657	1.652
Real Estate	1.646	1.713
Shopping	1.672	1.681
Sports	1.630	1.652
Travel	1.670	1.686

Table 2.2: RMSE results of Germany's model (19)

Country:	**Germany**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 and Google categories (19)		
Autos & Vehicles	1.130	1.164
Beauty & Fitness	1.162	1.163
Business & Industrial	1.154	1.170
Computers & Electronics	1.178	1.173
Food & Drink	1.163	1.187
Health	1.153	1.182
Home & Garden	1.086	1.172
Internet	1.179	1.181
Investing	1.171	1.170
Jobs	1.157	1.179
Law	1.168	1.193
News	1.162	1.089
Real Estate	1.123	1.157
Shopping	1.148	1.178
Sports	1.112	1.156
Travel	1.136	1.126

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 and Google categories (19)		
Autos & Vehicles	1.111	1.140
Beauty & Fitness	1.193	1.209
Business & Industrial	1.179	1.193
Computers & Electronics	1.353	1.259
Food & Drink	1.110	1.105
Health	1.168	1.462
Home & Garden	1.069	1.105
Internet	1.257	1.216
Investing	1.184	1.186
Jobs	1.194	1.202
Law	1.521	1.619
News	1.217	1.010
Real Estate	1.132	1.191
Shopping	1.225	1.175
Sports	1.112	1.148
Travel	1.130	1.093

Table 2.3: RMSE results of Finland's model (20)

Country:	**Finland**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 with Google and Confidence (20)		
Autos & Vehicles	1.673	1.647
Beauty & Fitness	1.771	1.714
Business & Industrial	1.712	1.645
Computers & Electronics	1.685	1.654
Food & Drink	1.756	1.707
Health	1.742	1.648
Home & Garden	1.686	1.636
Internet	1.707	1.661
Investing	1.735	1.636
Jobs	1.820	1.694
Law	1.714	1.660
News	1.766	1.754
Real Estate	1.683	1.636
Shopping	1.694	1.683
Sports	1.684	1.636
Travel	1.755	1.684

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 with Google and Confidence (20)		
Autos & Vehicles	1.673	1.687
Beauty & Fitness	1.846	1.776
Business & Industrial	1.887	1.680
Computers & Electronics	1.689	1.650
Food & Drink	1.723	1.683
Health	1.831	1.719
Home & Garden	1.709	1.673
Internet	1.721	1.674
Investing	1.728	1.616
Jobs	1.915	1.692
Law	1.696	1.669
News	1.692	1.618
Real Estate	1.711	1.655
Shopping	1.721	1.670
Sports	1.676	1.652
Travel	1.743	1.682

Table 2.4: RMSE results of Germany's model (20)

Country:	**Germany**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 with Google and Confidence (20)		
Autos & Vehicles	1.197	1.244
Beauty & Fitness	1.191	1.213
Business & Industrial	1.193	1.247
Computers & Electronics	1.228	1.257
Food & Drink	1.212	1.246
Health	1.193	1.251
Home & Garden	1.120	1.216
Internet	1.167	1.172
Investing	1.208	1.238
Jobs	1.223	1.280
Law	1.225	1.278
News	1.256	1.680
Real Estate	1.187	1.249
Shopping	1.184	1.243
Sports	1.204	1.253
Travel	1.188	1.208

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 with Google and Confidence (20)		
Autos & Vehicles	1.286	1.256
Beauty & Fitness	1.329	1.360
Business & Industrial	1.459	1.513
Computers & Electronics	1.675	1.577
Food & Drink	1.174	1.312
Health	1.538	2.077
Home & Garden	1.129	1.188
Internet	1.639	1.582
Investing	1.353	1.324
Jobs	1.267	1.450
Law	2.224	2.198
News	1.401	1.155
Real Estate	1.322	1.502
Shopping	1.387	1.306
Sports	1.203	1.210
Travel	1.203	1.225

APPENDIX 4 Dimension reduction before the nowcasting exercise

Following RMSE results and figures are for models where PCA and PLS dimension reduction methods this thesis conducted before the pseudo-out-of-sample exercise. In other words, these models used data that was ex-post available.

Table 3.1: Models that included entire Finland's Google Trends data

Country:	Finland	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google (17)	1.483	1.463
AR-1 and Google (19)	1.745	1.787
AR-1 with Google and Confidence (20)	1.800	1.703

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google (17)	1.472	1.458
AR-1 and Google (19)	1.734	1.770
AR-1 with Google and Confidence (20)	1.790	1.699

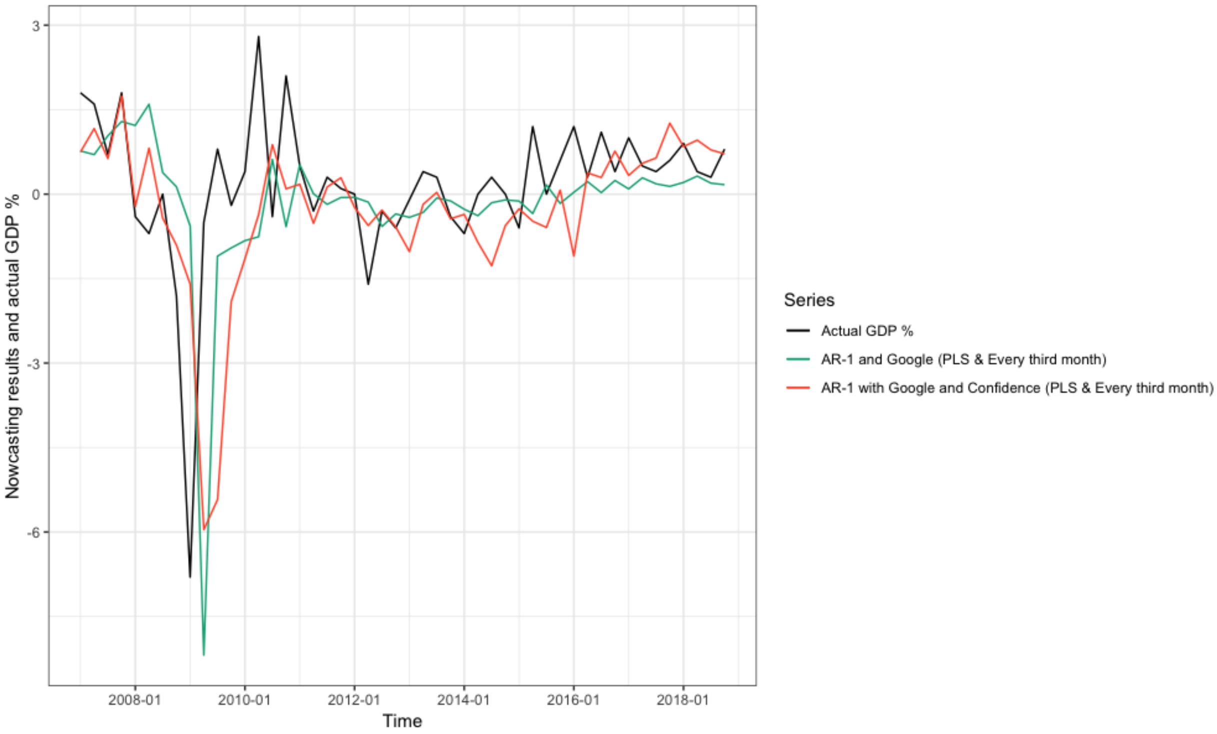


Figure 3.1: Google models and Finland's GDP growth

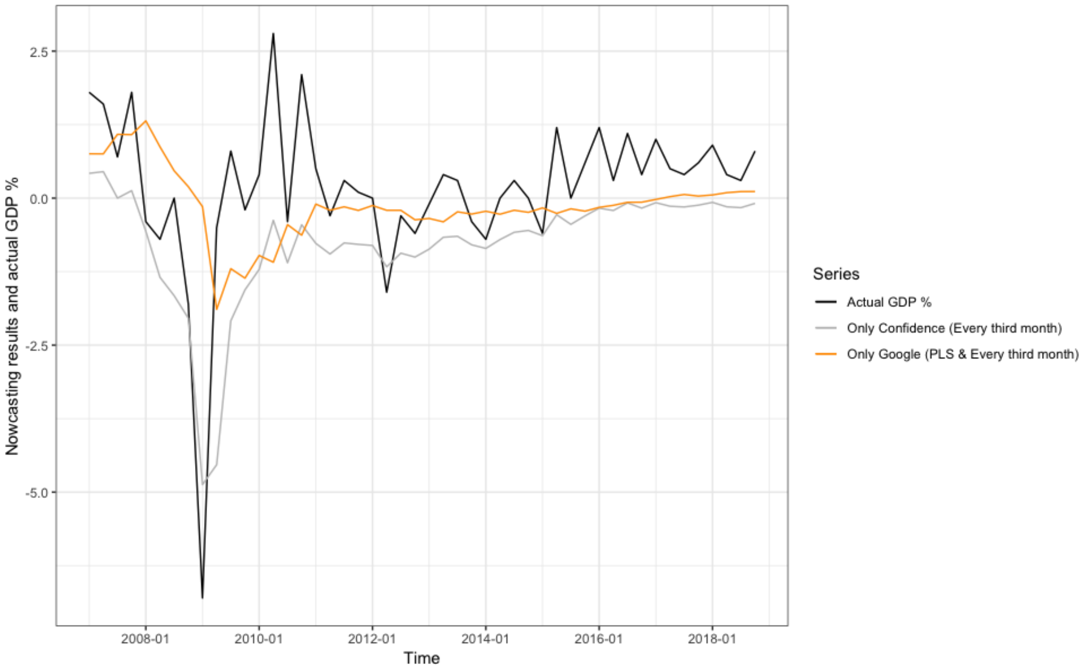


Figure 3.2: Leading Google model and Finland's GDP growth

Table 3.2: Models that included entire Germany's Google Trends data

Country:	**Germany**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google (17)	1.061	1.034
AR-1 and Google (19)	1.243	1.239
AR-1 with Google and Confidence (20)	1.235	1.239

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google (17)	0.968	0.970
AR-1 and Google (19)	1.238	1.241
AR-1 with Google and Confidence (20)	1.298	1.300

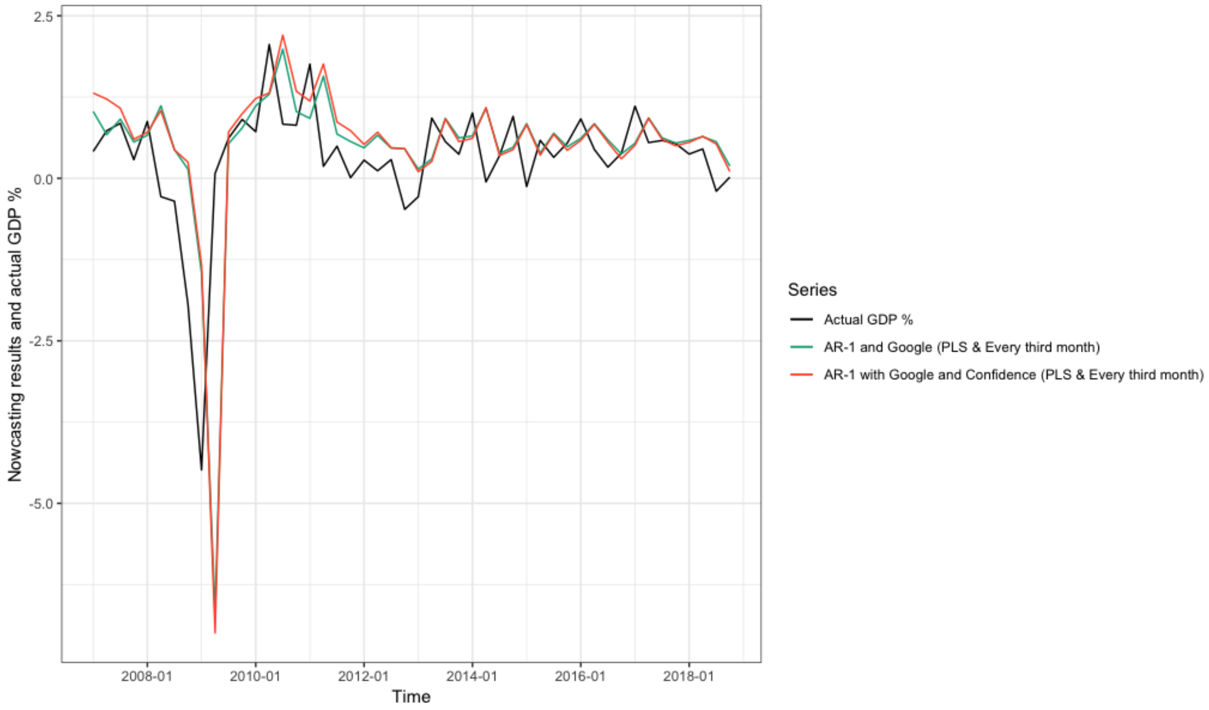


Figure 3.3: Google models and Germany's GDP growth

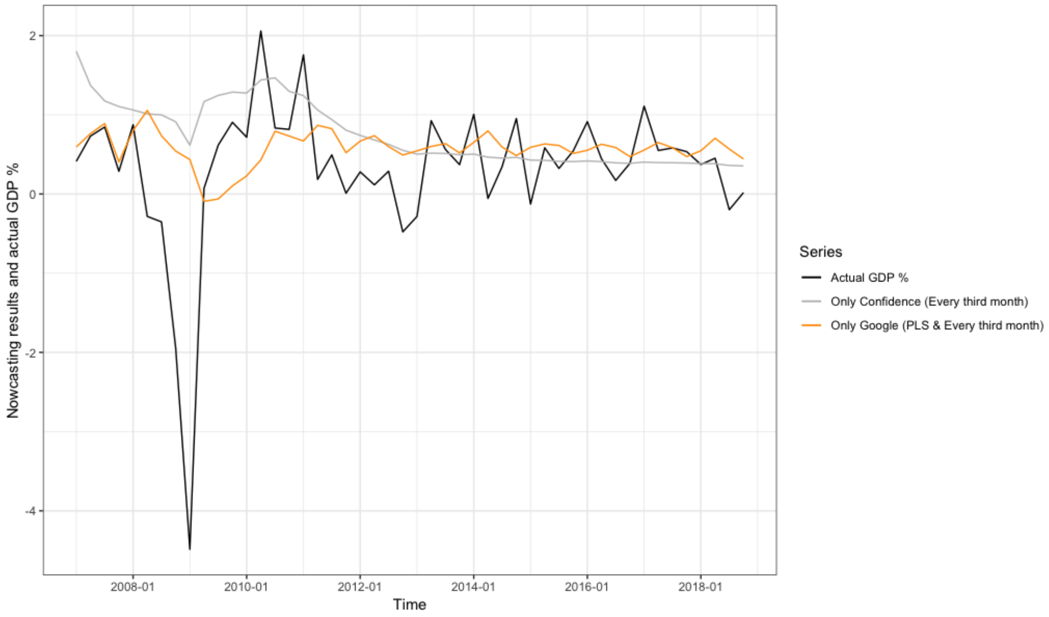


Figure 3.4: Leading Google model and Germany's GDP growth

Table 3.3: RMSE results of Finland's Google category models (17)

Country:	**Finland**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google categories (17)		
Autos & Vehicles	1.486	1.456
Beauty & Fitness	1.481	1.462
Business & Industrial	1.472	1.446
Computers & Electronics	1.450	1.443
Food & Drink	1.429	1.444
Health	1.458	1.446
Home & Garden	1.465	1.455
Internet	1.482	1.465
Investing	1.482	1.473
Jobs	1.443	1.425
Law	1.478	1.457
News	1.451	1.441
Real Estate	1.484	1.443
Shopping	1.482	1.461
Sports	1.484	1.470
Travel	1.463	1.447

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google categories (17)		
Autos & Vehicles	1.473	1.449
Beauty & Fitness	1.480	1.467
Business & Industrial	1.466	1.446
Computers & Electronics	1.449	1.444
Food & Drink	1.421	1.429
Health	1.457	1.444
Home & Garden	1.477	1.454
Internet	1.478	1.463
Investing	1.423	1.419
Jobs	1.442	1.425
Law	1.468	1.454
News	1.433	1.431
Real Estate	1.471	1.446
Shopping	1.477	1.462
Sports	1.498	1.485
Travel	1.461	1.445

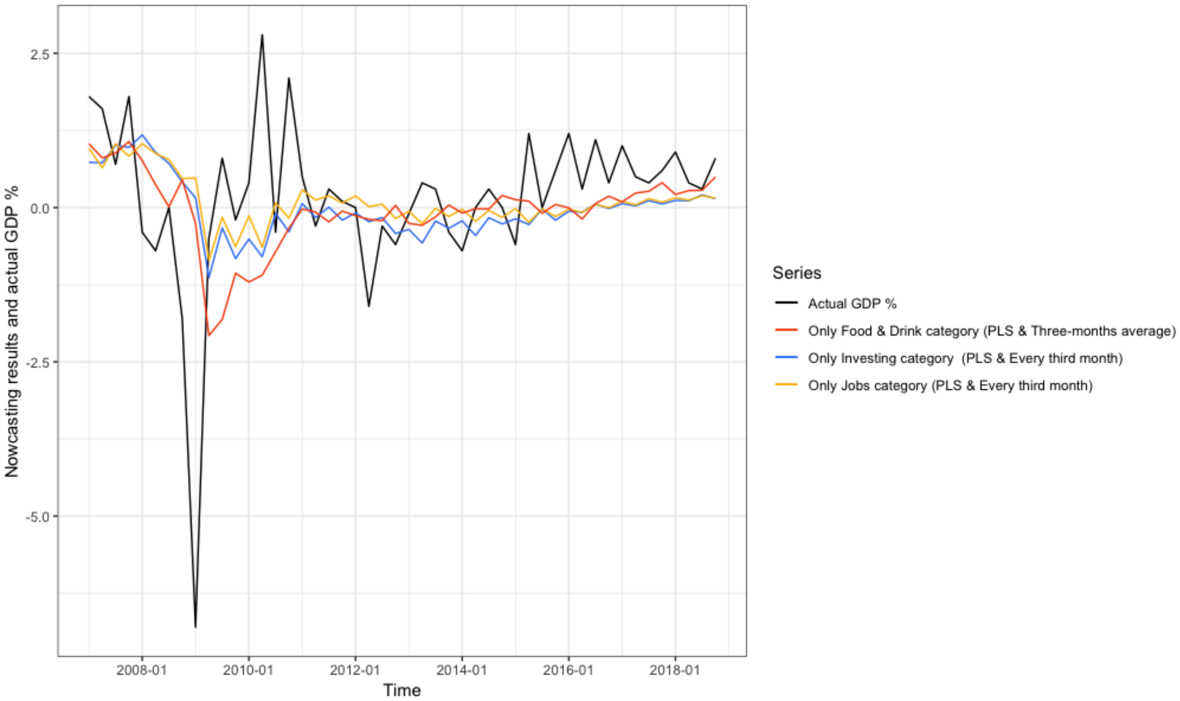


Figure 3.5: Leading Google category models and Finland's GDP growth

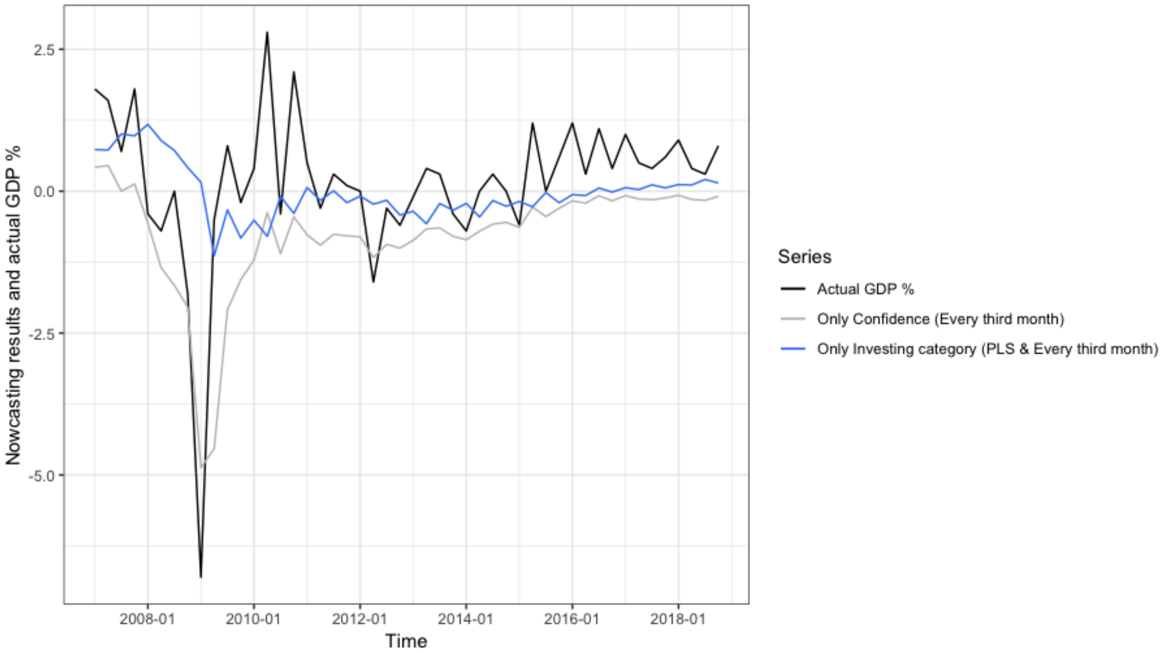


Figure 3.6: Confidence and the leading Google model against Finland's GDP growth

Table 3.4: RMSE results of Germany's Google category models (17)

Country:	**Germany**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google categories (17)		
Autos & Vehicles	1.027	1.008
Beauty & Fitness	1.058	1.043
Business & Industrial	1.049	1.019
Computers & Electronics	1.046	1.041
Food & Drink	1.065	1.031
Health	1.043	1.008
Home & Garden	1.072	1.055
Internet	1.087	1.078
Investing	1.062	1.037
Jobs	1.033	0.997
Law	1.040	1.007
News	1.001	1.004
Real Estate	1.037	1.014
Shopping	1.056	1.024
Sports	1.022	0.999
Travel	1.034	1.034

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
Only Google categories (17)		
Autos & Vehicles	0.958	0.916
Beauty & Fitness	1.063	1.053
Business & Industrial	1.035	1.009
Computers & Electronics	1.048	1.045
Food & Drink	1.004	1.004
Health	1.045	0.993
Home & Garden	0.956	0.947
Internet	1.006	1.015
Investing	1.043	1.031
Jobs	1.038	1.000
Law	1.036	0.998
News	0.967	0.926
Real Estate	1.056	1.001
Shopping	1.049	1.008
Sports	0.960	0.933
Travel	0.997	0.968

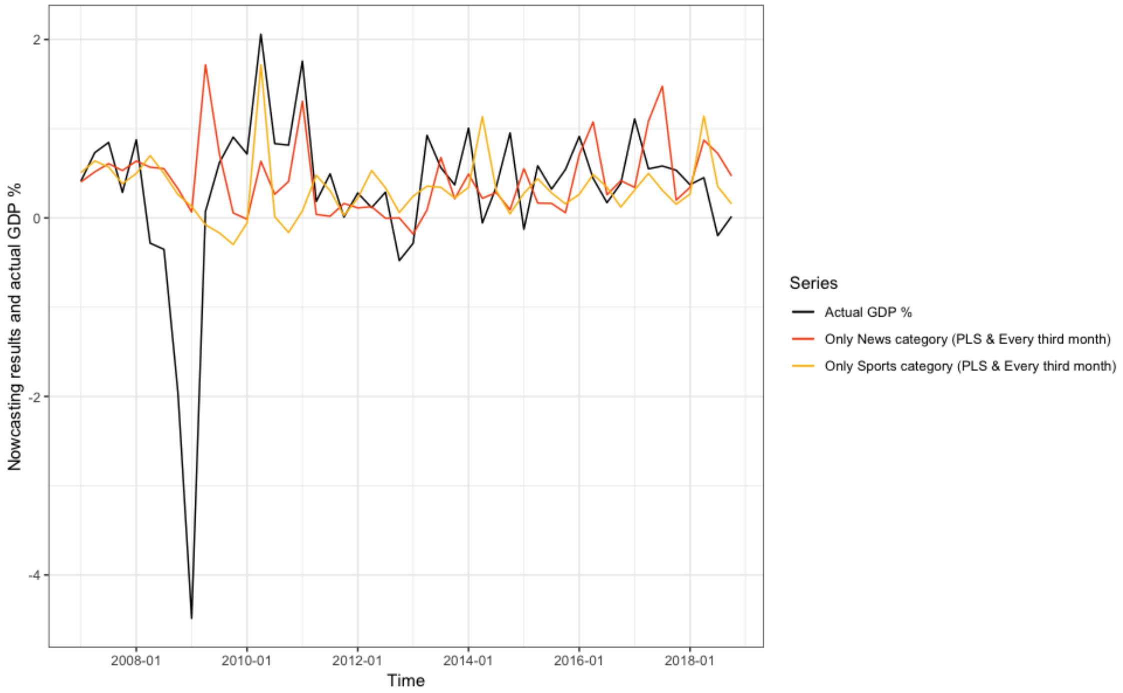


Figure 3.7: Two of the leading Google category models and Germany's GDP growth

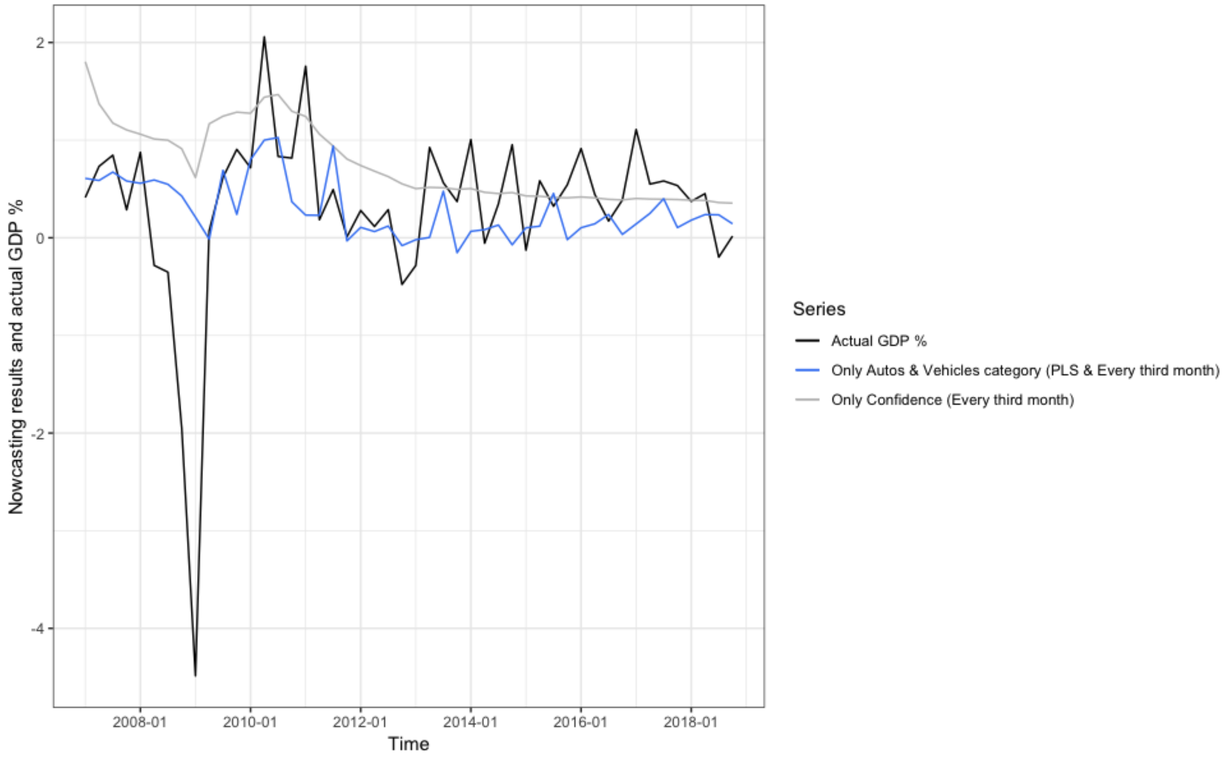


Figure 3.8: Confidence and the leading Google model against Germany's GDP growth

Table 3.5: RMSE results of Finland's model (19)

Country:	**Finland**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 and Google categories (19)		
Autos & Vehicles	1.734	1.759
Beauty & Fitness	1.771	1.805
Business & Industrial	1.755	1.799
Computers & Electronics	1.748	1.776
Food & Drink	1.744	1.815
Health	1.737	1.807
Home & Garden	1.737	1.765
Internet	1.752	1.797
Investing	1.765	1.815
Jobs	1.783	1.821
Law	1.736	1.800
News	1.809	1.883
Real Estate	1.734	1.788
Shopping	1.769	1.798
Sports	1.752	1.754
Travel	1.742	1.747

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 and Google categories (19)		
Autos & Vehicles	1.713	1.747
Beauty & Fitness	1.765	1.790
Business & Industrial	1.749	1.792
Computers & Electronics	1.749	1.775
Food & Drink	1.734	1.773
Health	1.736	1.801
Home & Garden	1.730	1.761
Internet	1.748	1.786
Investing	1.750	1.775
Jobs	1.783	1.820
Law	1.733	1.779
News	1.756	1.777
Real Estate	1.724	1.781
Shopping	1.748	1.772
Sports	1.748	1.744
Travel	1.738	1.737

Table 3.6: RMSE results of Germany's model (19)

Country:	**Germany**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 and Google categories (19)		
Autos & Vehicles	1.215	1.226
Beauty & Fitness	1.258	1.254
Business & Industrial	1.253	1.247
Computers & Electronics	1.276	1.267
Food & Drink	1.228	1.221
Health	1.250	1.250
Home & Garden	1.255	1.250
Internet	1.260	1.240
Investing	1.258	1.271
Jobs	1.255	1.238
Law	1.263	1.263
News	1.226	1.193
Real Estate	1.235	1.227
Shopping	1.241	1.232
Sports	1.193	1.207
Travel	1.217	1.203

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 and Google categories (19)		
Autos & Vehicles	1.168	1.193
Beauty & Fitness	1.308	1.283
Business & Industrial	1.281	1.268
Computers & Electronics	1.277	1.266
Food & Drink	1.205	1.228
Health	1.272	1.283
Home & Garden	1.185	1.201
Internet	1.234	1.232
Investing	1.252	1.295
Jobs	1.256	1.237
Law	1.269	1.286
News	1.276	1.129
Real Estate	1.257	1.247
Shopping	1.236	1.229
Sports	1.177	1.186
Travel	1.223	1.210

Table 3.7: RMSE results of Finland's model (20)

Country:	**Finland**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 with Google and Confidence (20)		
Autos & Vehicles	1.792	1.689
Beauty & Fitness	1.813	1.706
Business & Industrial	1.796	1.701
Computers & Electronics	1.779	1.695
Food & Drink	1.791	1.723
Health	1.799	1.709
Home & Garden	1.795	1.698
Internet	1.792	1.710
Investing	1.814	1.715
Jobs	1.804	1.701
Law	1.800	1.707
News	1.806	1.755
Real Estate	1.797	1.693
Shopping	1.803	1.712
Sports	1.797	1.695
Travel	1.798	1.690

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 with Google and Confidence (20)		
Autos & Vehicles	1.784	1.689
Beauty & Fitness	1.803	1.703
Business & Industrial	1.794	1.700
Computers & Electronics	1.779	1.695
Food & Drink	1.784	1.711
Health	1.796	1.706
Home & Garden	1.794	1.695
Internet	1.786	1.707
Investing	1.783	1.689
Jobs	1.802	1.701
Law	1.788	1.701
News	1.785	1.708
Real Estate	1.792	1.699
Shopping	1.790	1.711
Sports	1.797	1.703
Travel	1.795	1.689

Table 3.8: RMSE results of Germany's model (20)

Country:	**Germany**	
Dimension reduction method:	**Principal component analysis (PCA)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 with Google and Confidence (20)		
Autos & Vehicles	1.207	1.240
Beauty & Fitness	1.266	1.269
Business & Industrial	1.238	1.242
Computers & Electronics	1.263	1.261
Food & Drink	1.268	1.277
Health	1.250	1.263
Home & Garden	1.274	1.271
Internet	1.292	1.270
Investing	1.290	1.293
Jobs	1.253	1.247
Law	1.261	1.273
News	1.225	1.166
Real Estate	1.255	1.252
Shopping	1.269	1.278
Sports	1.193	1.226
Travel	1.219	1.198

Dimension reduction method:	**Partial least squares (PLS)**	
Models:	**RMSE Results:**	
	Three-months average	**Every third-month**
AR-1 with Google and Confidence (20)		
Autos & Vehicles	1.153	1.243
Beauty & Fitness	1.294	1.289
Business & Industrial	1.262	1.255
Computers & Electronics	1.268	1.268
Food & Drink	1.200	1.289
Health	1.271	1.307
Home & Garden	1.246	1.260
Internet	1.302	1.320
Investing	1.300	1.343
Jobs	1.253	1.246
Law	1.275	1.322
News	1.328	1.196
Real Estate	1.292	1.294
Shopping	1.281	1.288
Sports	1.226	1.237
Travel	1.235	1.240