ABSTRACT

This study conducts a comprehensive investigation into the forecasting of private consumption expenditure (PCE) in the United States while considering the transformative impact of the COVID-19 pandemic. PCE, a pivotal driver of the nation's GDP, demands meticulous prediction for informed decision-making. The primary objective is to assess the effectiveness of various economic indicators, including Google Trends data, survey-based indices such as the Michigan Consumer Sentiment Index (MCSI) and the Consumer Confidence Index (CCI), and the Economic Policy Uncertainty (EPU) index. Employing the ARIMAX model, the study spans two crucial periods: the pre-COVID era (2005-2019) and the pandemic era (2005-2023). In the pre-COVID landscape, survey-based indices exerted notable influence in predicting PCE across various timelines. However, as the pandemic unfolded, Google Trends emerged as a formidable predictor for all components of consumption, particularly for durable goods and services reflecting a digital transformation. In times of economic turbulence, the Economic Policy Uncertainty (EPU) index gained prominence, particularly impacting overall consumption and durable goods. Notably, the fusion of Google Trends and the EPU index as well as the combined model, which integrates all indicators, demonstrated exceptional forecasting capabilities during the COVID-19 period. These findings offer valuable insights for policymakers and businesses navigating with economic uncertainty and evolving consumer behaviours[1].

Keywords: Google Trends, Michigan Consumer Sentiment Index, Consumer Confidence Index, Economic Policy Uncertainty, private consumption expenditure

[1] This dissertation follows the structure used by Vosen & Schmidt (2011) and Woo and Owen (2018)

ACKNOWLEDGMENT

I wish to extend my heartfelt appreciation to Dr. Panagiotis Stamolampros, whose unwavering support and mentorship have been instrumental in shaping the course of my research journey. His guidance and expertise have been invaluable. I would also like to express my gratitude to the dedicated faculty members of LUBS whose wisdom and guidance have greatly enriched my academic pursuit.

My time at the University of Leeds has been made more meaningful by the friendships I have cultivated here. I extend my warm thanks to my friends whose companionship has transformed this postgraduate experience into a truly remarkable and unforgettable chapter of my life.

Lastly, but by no means least, I owe a debt of gratitude to my family and to the Divine. Their unwavering support and faith in me have been my pillars of strength throughout this challenging expedition.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Forecasting Private Consumption Expenditure Using Google Trends, Survey-Based Indicators and Economic Policy Uncertainty Index: Analysing the Impact of COVID-19 in USA

1. INTRODUCTION

Household consumption expenditure is a crucial component of overall demand in an economy, and typically makes up around 70% of the United States' GDP. Household consumption expenditure, also known as private consumption expenditure (PCE), refers to the total value of goods and services purchased by households, including durable products such as cars and appliances, as well as payments made to governments for permits and licenses. It does not include purchases of homes, but it does include imputed rent for homes that are owner-occupied (Bonsu and Muzindutsi, 2017). Business officials and government pay close attention to shifts in consumer spending, making it crucial to accurately forecast overall consumer spending patterns.

Macro-economic indicators are statistics that provide information about the overall performance of an economy. The Michigan Consumer Sentiment Index (MCSI) is a monthly survey conducted by the University of Michigan that measures the level of confidence that U.S. consumers have in the overall state of the economy and is used to predict aggregate trends in consumer expenditures (Curtin, 1982). The survey is conducted by asking a representative sample of consumers about their current financial situation and their expectations for the economy in the future. Conference Board Consumer Confidence Index (CCI) is another survey-based indicator which measures similar factors with similar methodologies as that of MCSI, yet it emphasises more on the health of the labour market and unemployment rate (Vosen & Schmidt,2011). Carroll et al. (1994) and Ludvigson (2004) proved that these consumer sentiment indicators have significant power in predicting the US consumption over other macroeconomic variables.

The Economic Policy Uncertainty (EPU) index is created to quantify economic uncertainty and is based on three separate factors to assess uncertainty: press coverage, stock market volatility, and expectations as discovered through company surveys. The indeterminacy of economic and financial decisions under increased EPU could result in changes of government policies. It is difficult for individuals, households, and corporations to make appropriate economic decisions without proper, transparent economic policies (Al-Thaqeb, Algharabali and Alabdulghafour, 2020). Therefore, it is important to consider this economic indicator for forecasting private consumption expenditure.

According to statistics, 81% of consumers conduct online research before buying a product and on an average 1.62% of ecommerce website visits converts into a purchase (Law, 2019). The availability of real time data, large sample size, granularity of data allowing people to search for specific key terms and free accessibility makes Google trends a valuable resource for research and business. The most alluring factor about Google trends is how it maintains the data quality by eliminating duplicates, special characters, avoiding searches made by very few people and anonymises the identity of people (Choi and Varian, 2012).

"The macroeconomic variables indicate the consumers' ability to spend, survey-based indicators represent their willingness to spend whereas the Google trends indicator provides a measure for consumers preparatory steps towards spending" (Vosen & Schmidt,2011).

This study examines the use of various economic indicators to enhance private consumption expenditure (PCE) prediction in the United States. The analysis utilizes MCSI, CCI, Google Trends and Economic Policy Uncertainty (EPU) index, categorizing consumption according to the Bureau of Economic Analysis (BEA) categories (Bea.gov,2017). Historical data from google trends search categories and other economic indicators are incorporated into the models alongside lagged data of the total PCE and its components. Forecasting the private consumption expenditure for USA is done using the ARIMAX model.

1.1.<u>Research Objective</u>

Analyse the impact of economic indicators in forecasting private consumption expenditure in USA, focusing on Google Trends, survey-based indicators (MCSI, CCI), and the Economic Policy Uncertainty (EPU) index, while considering the disruptive influence of the COVID-19 pandemic.

1.2.<u>Research Questions</u>

This research aims to address the following questions:

i. How does the incorporation of Google Trends data, survey-based indicators (MCSI and CCI), and the EPU index impact the accuracy of forecasting private consumption expenditure using the ARIMAX model?
ii. How have consumption patterns evolved from 2005 to 2023, considering major events like the Great Recession and COVID-19? What are the trends, shifts in behaviour, and the impact of the COVID-19 crisis on durable goods, non-durable goods, and services consumption?
iii. What is the significance of economic policy uncertainty (EPU) indicator in predicting variations in private consumption expenditure? How do these policy uncertainties interact with consumer sentiment and online search behaviour captured by Google Trends data?

iv. What is the combined effect of Google Trends data, survey-based indicators, and the EPU index on the accuracy of forecasting private consumption expenditure, and how do these variables interact with each other in influencing consumer spending patterns?

This paper is structured in the following manner. Section 2 delves into a comprehensive literature review regarding the significance of survey-based indicators, google trends and economic policy uncertainty index in driving the economy, their utility in predicting private consumption expenditure and significance of ARIMAX model. Section 3 outlines the data collection, data transformation and methodology employed. Section 4 unveils the results and discussion with supporting evidence based on previous literature. Section 5 concludes the study answering the research questions, practical implications, limitations, and future research.

2. LITERATURE REVIEW

Forecasting is a vital aspect of business decision-making, acting as a bridge between goals and actionable strategies. It analyses historical data and potential events, offering valuable insights for both short-term scheduling and long-term strategic decisions. This comprehensive understanding of future demands, opportunities, and challenges empowers organizations to make informed decisions. Successful forecasting systems necessitate expertise, appropriate methods, continuous evaluation, and robust organizational support for implementation. (Hyndman and Athanasopoulos, 2018).

Effective forecasting relies on data availability and relevance. Qualitative methods come into play when historical data is scarce or unsuitable, providing structured approaches for accurate forecasts. Conversely, quantitative methods are employed when past numerical data is available, and continuity in patterns is reasonably assumed. These methods encompass various techniques tailored to specific disciplines and purposes. Among them, time series forecasting is pivotal, concentrating on sequentially observed data over time. It enables businesses to anticipate future developments and the evolution of sequential observations in the days to come. (Hyndman and Athanasopoulos, 2018).

Ingle et al (2021) conducted a review of demand forecasting methodologies across industries, categorizing them into traditional statistical, machine learning, deep learning, and hybrid models, and emphasized the importance of selecting models based on dataset characteristics and problem requirements rather than solely relying on advanced techniques. Univariate datasets perform better with traditional statistical models, while complex and unstable datasets are suited for machine learning or deep learning models. Another study by Krollner, Vanstone and Finnie (2010) examined the use of machine learning methods and artificial intelligence, in forecasting stock market movements.

Forecasting is an important tool for predicting macroeconomic indicators. The main variables forecasters try to pin down include output growth, inflation, unemployment, interest rates, exchange rates, international trade flows, fiscal balances and public debt (N.Carnot et al., 2005). Fulton and Hubrich (2021) examined real-time forecasting of US inflation and shows how incorporating additional information, like macroeconomic variables and expert judgment, enhances accuracy and resilience. Their study asserts that simple models remain competitive, but forecast combination, aggregating inflation components, and using judgmental forecasts improve overall performance, especially after the Global Financial Crisis. Zhang et al. (2019) proposed a novel method for product sales forecasting highlighting the significance of incorporating macroeconomic indicators and online reviews with sentiment analysis based on prospect theory.

In the realm of economic forecasting, consumption holds significant importance. While various factors like business, government, and consumer financial activities contribute to national economies, consumer spending serves as the primary driving force behind the U.S. economy. In the U.S., household consumption forecasting primarily relies on survey-based indicators, offering valuable insights into future private consumption behaviours (Chron,2020). The following sections will discuss the significance of survey-based indicators in driving the economy and its contribution to forecasting private consumption expenditure.

The COVID-19 pandemic has had a profound impact on household consumption patterns, with significant shifts observed in spending behaviours globally. Baker et al. (2020) analysed household consumption patterns during the COVID-19 pandemic in the United States, revealing that households underwent significant changes in their spending habits as the number of COVID-19 cases increased. Initially, there was a sharp increase in spending, particularly in retail, credit card purchases, and food items. However, this was followed by a notable decrease in overall spending. Households in states with shelter-in-place orders implemented by March 29th showed the most pronounced response in altering their spending behaviour. The impact of social distancing measures was particularly evident in reduced spending on restaurants and retail. According to Cavallo (2020), as a result of lockdowns and social-distancing measures, consumer spending shifted towards certain categories like "Food and Beverages," which experienced more inflation, and away from categories like "Transportation," which witnessed significant deflation. The research demonstrates that the COVID-19 inflation rate is higher than the official CPI in the US and 10 out of 16 additional countries. The study reveals that low-income households experience higher inflation during the crisis compared to high-income households. These findings have important implications for policymakers, as they suggest that the cost of living for consumers is higher than what is implied by the official CPI.
Hence, considering the far-reaching effects of COVID-19 on consumer behaviour, this project aims to analyse its impact on private consumption expenditure patterns in-depth.

2.1. Significance of Survey Based Indicators on the Economy

An economic growth is commonly defined as an expansion in the economy's capacity to produce goods and services over time. In the field of macroeconomics, this growth is typically measured by changes in real Gross Domestic Product (GDP). Kim (2016) studied the relationship between consumer sentiment and GDP in ten countries. It found that changes in consumer sentiment, measured by the Consumer Sentiment Index (CSI), can be a leading indicator of economic movements, with increased sentiment often preceding economic growth. Reverse causality was also observed, indicating that economic growth can boost consumer confidence. The Consumer Confidence Index (CCI) reflects people's economic confidence and influences saving and spending habits, contributing to economic expansion. Tanweer UI Islam (2016) found a long-term relationship between CCI and economic growth in European Union countries like United Kingdom, Germany, France, Denmark, and Netherlands. The research proved that positive changes in consumer confidence will lead to the economic growth of countries and vice versa. Mazurek and Mielcová (2017) studied the predictive relationship between the Consumer Confidence Index (CCI) and real GDP growth in the USA during economic recessions (1960-2015). They found that CCI can serve as a suitable predictor for US GDP, using the Engle-Granger test for long-term dependency and Granger causality for short-term relations.

The collective evidence from these studies highlights the significant impact of survey-based indicators on the economic activity of a country. The effectiveness of these indicators transcends national boundaries, making them a crucial tool for policymakers and researchers seeking to understand and forecast economic trends on a global scale.

2.2. Significance of Survey-Based Indicators in forecasting PCE

Increasing consumer demand for goods and services is a crucial factor driving the economic growth of the United States. Consumer spending, as measured by personal consumption expenditures (PCE) accounts for 67.9% of the gross domestic product of United States in March,2023 (CEIC, 2018). According to the U.S Bureau of Labour Statistics, consumer spending helps business in determining the value of their products in the economic marketplace.

Consumer spending reflects the overall confidence of consumers in a country’s economy. The University of Michigan’s Consumer Sentiment Index and the Conference Board’s Consumer Confidence Index are used to measure the US consumer confidence. Both the surveys, base their overall consumer confidence index on five questions covering various aspects including current economic conditions (financial situation, job prospect and overall economic conditions), economic outlook (respondent’s expectations for the future business conditions, employment opportunities), consumer spending (buying goods, future spending plans), income, savings, and demographics. Michigan conducts monthly phone surveys on a sample of 500 people, whereas the Conference board sends monthly mail surveys to 5000 people (Ludvigson,2004).

The survey responses are consolidated into a single number, which serves as an "index" representing consumer confidence. Consumer confidence indexes are important because they can provide an early indication of the economy's strength due to two main reasons. First, there is a correlation between consumer confidence and current economic conditions, which could be because consumers accurately reflect current conditions or because their confidence influences their spending behaviour. Second, consumer confidence indexes can provide valuable forecasts of future economic activity, either because consumer confidence has a direct impact on economic activity or because consumers are skilled at predicting the economy. Hence, consumer confidence serves as a convenient summary for forecasting individual information (Piger,2003).

The predictive power of survey-based consumer sentiment indicators for private consumption have been proven in many academic literatures. Carroll, Fuhrer and Wilcox (1994) proved that if consumer sentiment is a part of uncertainty, then consumption growth is negatively correlated with contemporaneous uncertainty, but positively correlated with lagged uncertainty. Mishkin et al. (1978) explored the impact of the consumer sentiment index (ICS) on the purchase of durable goods. They found that consumer sentiment reflects consumer perceptions of the likelihood of financial distress, which can strongly influence the decision to purchase durables due to their illiquid nature. Eppright et.al (1998) used multivariate vector autoregression (MVAR) analyses to investigate the predictive power of aggregate consumer attitude and expectation measures, such as the University of Michigan Index of Consumer Sentiment and the Conference Board Index of Consumer Confidence. The results showed that these measures were more successful in predicting future aggregate consumer spending than economic indicators, suggesting that they may be useful tools for anticipating future trends in consumer behaviour. Howrey (2001) considered four recession indicator series: the University of Michigan Index of Consumer Sentiment (ICS), long-term and short-term interest rates, the New York Stock Exchange composite price index and the Conference Board index, both by itself and in combination. The monthly analysis proved the ICS index to be both statistically and economically significant in terms of point forecasts of the rate of growth of personal consumption expenditure. The quarterly analysis of both lagged and current quarter monthly values of ICS are statistically significant but only a slight reduction of standard error of forecast of consumption expenditure for the quarter.

Bram and Ludvigson (1998) suggested that consumer confidence can help predict consumption and consumer attitudes can also act as a catalyst to economic fluctuations. They compared the forecasting power of MCSI and CCI and found that lagged values of CCI provide more information about the future consumer spending that is not captured by MCSI and other macro-economic indicators. Laaksoharju (2011) investigated the correlation and causality between consumer sentiments like MCSI, CCI and stock markets data (SP500) in USA using Pearson-R and Granger methods to conclude that there exists a relationship and the indices move in the same direction in the long run, with CCI showing a slightly stronger relationship.

Despite these studies preferring CCI over MCSI, Mowen, Young and Silpakit (1985) stated that the MCSI and CCI measures different concepts. The Index of Consumer Sentiment primarily evaluates consumer reactions to financial factors, including prices, interest rates, the stock market, and the length of the work week. In contrast, the Index of Consumer Confidence captures additional information on employment-related variables, such as the length of the work week, accession-layoff rate, and disposable income. In essence, the University of Michigan’s Sentiment index predominantly assesses consumer perceptions of their future financial assets, while the Conference Board Index predominantly measures consumer perceptions of their employment prospects.

2.3. Significance of Google Trends in forecasting PCE

According to recent estimates, the global big data market is projected to reach a revenue of $4273.4 billion by 2026. This growth is driven by a significant increase in the volume of data being generated and collected. As a result, businesses across various industries are implementing big data solutions to effectively manage and analyze this data, to gain valuable insights and make informed decisions (Marketsandmarkets, 2016). Google Trends is a free online service, launched by Google LLC in 2008, that enables users worldwide to examine and analyse big data. Gingsberg et al. (2009) published a pioneeering study forecasting the spread of influenza using Google Trends ahead of the Centers for Disease control and Prevention. Shim et al. (2001) proved the positive correlation between consumers online search and product purchase, which was further justified by To et al. (2007).

The first published research stating the effectiveness of web search data in economic forecasts was by Ettredge et al.(2005) which examined US unemployment rate and stated that search engine data can be used to identify people's needs, wants, interests and conserns. Choi and Varian (2009, a, b) demonstrated the use of Google trends insights data to forecast economic metrics like early unemployment claims and sales of retail, automotives, home and travel. Followed by this research they also demonstrated the use of search engine data to forecast short-term economic indices (Choi and Varian,2012). According to their research, Google Trends has a "now-casting" function and has the potential to become a powerful forecasting tool in various fields. Penna and Huang (2009) showed that Google Trends, MCSI and CCI have 90% correlation. The most interesting finding of this research is that Google Trends can be used to predict the survey-based indices, but changes in MCSI or CCI cannot be used to predict the Google trend index. Hence, the study infers that these three indicators are compliments of each other.

The relationship between consumption related Google Trends data and private consumption expenditure was first examined by Vosen & Schmidt (2011) using AutoRegressive Moving Average (ARMA) model which proved Google trends index to be a better predictor for private consumption than MCSI and CCI.

Fasulo et al. (2017) used ARIMA model to demonstrate the ability of Google Trends to nowcast quarterly household consumption expenditure in Italy. They tested the model including leading indicators correlated to consumption like GDP, interest rate of Treasury Bills and Consumer Confidence Index on a data from 2004-2016, in which Google data outperformed the benchmark and augmented models. Gil et al. (2018) focused on nowcasting and forecasting quarterly private consumption using a mix of traditional indicators, uncertainty measures, credit card data, and internet-based indicators. The results reveal that while traditional indicators perform well individually, combining them with novel data sources, such as payment card transactions, Google-based data, and uncertainty indicators, enhances forecasting accuracy, particularly beyond the nowcasting horizon.

Woo & Owen (2018) proved the predictive relationship between private consumption in USA with that of consumption related and news related Google Trends data. Their results indicates that the consumption related data provide information about pre-consumption trends, the news related data provide information about changes in durable goods consumption and the combination of the two improves the performance of forecasting models when compared to survey-based indicators.

2.4.Significance of Economic Policy Uncertainty in forecasting PCE

Following the financial crisis in the world, the crisis in the Eurozone, and the party strife in the United States, there has been a noticeable increase in policy uncertainty. This uncertainty contributed to economic downturn of 2008 and slow economic recovery. Similarly, events like the "Arab Spring," which caused political unrest in the Middle East, and Donald Trump's election as president of the United States, which supported changes to the global status quo, have exacerbated political and economic volatility worldwide. Additionally, events like the Brexit vote in the UK and Russia's annexation of Crimea, have strained relations around the world and increased uncertainty (Al-Thaqeb, Algharabali and Alabdulghafour, 2020).

Overall, the current rise in uncertainty is largely caused by political conflict, polarisation, and the rising importance of government spending in the economy (Baker et al.,2013)

In the light of these events, the Economic Policy Uncertainty (EPU) Index was first introduced by Baker et al. (2013) by combining three components: news coverage, tax code expiration data, and economic forecaster disagreement. The news coverage component is an index of search results from 10 leading U.S. newspapers, that entails examining the frequency of terms like "economic," "uncertain," and policy-uncertainty. The tax code expiration data component draws on reports by the Congressional Budget Office (CBO) that compile lists of temporary federal tax code provisions, which can be a source of uncertainty for businesses and households. The economic forecaster disagreement component measures dispersion in the individual-level data for three forecast variables directly influenced by government policy: CPI, purchases of goods and services by state and local governments, and purchases of goods and services by the federal government. Each component is normalized by its own standard deviation prior to January 2012, and the overall index is computed as the average value of the components, with weights of 1/2 on the news-based policy uncertainty index and 1/6 on each of the other three measures. Their study highlights that key events like presidential elections, protracted wars, and economic crises spikes the EPU index. After the financial crisis of 2008, it significantly increased. A rise in policy uncertainty may result in a fall in GDP and employment. The study emphasises how policy ambiguity affects economic decision-making.

Binder (2016) explores the relationship between household inflation uncertainty and economic policy uncertainty. In comparison to expert forecasters, the research shows a higher association between policy uncertainty and consumer inflation uncertainty. Consumers with higher incomes and education levels are most affected by this connection since their expectations affect inflation dynamics. The study also reveals a bidirectional Granger causal relationship between long-term inflation uncertainty and policy uncertainty. Notably, customers with high incomes are least affected by policy uncertainty, potentially because of successful Federal Reserve communication strategies. The research extends to other countries, indicating U.S. policy uncertainty influences inflation uncertainty abroad. Balcilar, Gupta, and Segnon (2016) highlights the crucial role of EPU as a leading indicator in predicting movements in the quarterly GDP growth rates and associated recessionary periods within the U.S. economy by utilizing a mixed-frequency Markov-switching vector autoregressive (MF-MS-VAR) model that provides a better fit for different recession regimes.

Bloom (2009) reveals that uncertainty shocks prompt temporary fluctuations in overall output, employment, and productivity due to firms pausing their investment and hiring activities. Neglecting capital adjustment costs introduces biases, stressing the significance of factoring in both labour and capital adjustment costs. Leduc and Liu (2016) explore the connection between consumer expenditure and uncertainty concluding that increased uncertainty reduces aggregate demand, which results in lower consumer expenditure. Frictions in the job market and nominal rigidities strengthen this effect. Because of the uncertainty, businesses post fewer openings, which lowers job-finding rates and raises unemployment. Household income declines as a result, further reducing consumer spending and enhancing the recessionary effects of uncertainty. Using a Smooth Transition VAR model with post-World War II data, Caggiano, Castelnuovo, and Figueres (2017) examined the relationship between economic policy uncertainty (EPU) and unemployment in the United States. Their research shows that an unexpected rise in EPU has a more noticeable influence on unemployment during recessions than during expansions, showing that the impact of uncertainty differs depending on the economic environment. The study emphasises how crucial it is to appreciate the nonlinear dynamics between EPU and unemployment to fully comprehend the actions of the labour market and changes in the business cycle. Phan et. Al (2020) reveals a substantial negative impact of EPU on financial stability, showcasing the disruptive effect of uncertainty on information flow, liquidity, and market dynamics within the financial system. The findings underscore the relevance of understanding and managing EPU to ensure the resilience and stability of financial markets.

2.5. Using Time Series Model for Forecasting PCE using ARIMAX

Economic forecasting employs four main categories of quantitative approaches: subjective methods, indicator-based methods, time series models, and structural macroeconomic models. Time series models are solely based on the statistical properties of the series under consideration, irrespective of any interpretation or causal relationships informed by economic theory. Time series methods offer clear advantages: they are simple, require few data, and can be relatively successful even if the analyst knows little about the phenomenon under consideration (N.Carnot et al. 2005).

The Box Jenkins Model is a technique designed to forecast data ranges based on inputs from a specified time series. The ARIMA model is a form of Box Jenkins Model which consists of three components: autoregression (AR) for lagged observations, integration (I) for differencing to achieve stationarity, and moving average (MA) for residual errors from a moving average model. The parameters p, d and q determine the behaviour of these components. ARIMAX (AutoRegressive Integrated Moving Average with eXogenous variables) is a dynamic regression model that incorporates external variables into the ARIMA model, which can improve forecasting accuracy by capturing additional information that may affect the dependent variable.

Peter and Salvia (2012) use ARIMA and ARIMAX to analyse macro-economic time series. The R-Squared statistics of the ARIMAX fitted model explained 92.7% of the variance in the gross domestic product per capita. Fajar et al. (2020) utilized ARIMAX modelling to forecast Indonesia's unemployment rate during the COVID-19 pandemic by incorporating Google Trends data. The model demonstrates good forecasting capabilities with a MAPE value of 13.46%. By leveraging real-time Google Trends data, the model effectively captures current economic conditions and reflects the causal relationship between variables for accurate predictions. Another study on health expenditure in South Africa used ARIMAX to prove that health service costs and household disposable income had significant impact on household’s final consumption expenditure on health services (Munyaradzi Ganyaupfu, 2020).

The study by Serafini et al. (2020) focused on sentiment-driven price prediction for Bitcoin using statistical (ARIMAX) and deep learning (RNN) approaches. It highlighted the superior power of ARIMAX over the deep learning model in capturing sentiment-related patterns and providing more accurate forecasts. Abdulazeez and Adeyinka (2021) compared ARIMA and ARIMAX models for forecasting gross domestic product in Nigeria, where ARIMAX performed better than ARIMA. Andreas et al. (2022) compared ARIMAX and VARX models for cash flow forecasting in the banking sector during the COVID-19 pandemic. ARIMAX showed higher accuracy in predicting cash flows in the office and cash out of the office, while VARX performed better in predicting e-channel cash inflows. The study suggests that ARIMAX can be used to support cash flow balance in the banking sector during economic challenges.

2.6. Addressing the research gap

This research builds upon the groundwork laid by Woo & Owen (2018) and Vosen & Schmidt (2011) but extends their analysis beyond the pre-COVID-19 era. We expand the timeline to encompass the years from 2004 to 2023, providing insights into the repercussions of the economic crisis, including the profound influence of the COVID-19 pandemic. This extended timeframe also includes a critical juncture—the Great Recession of 2008. While prior studies offered limited exploration of the effectiveness of ARIMAX time series model in predicting private consumption expenditure, this study seeks to bridge that gap. The time series models augmented with Google Trends, MCSI, CCI and economic policy uncertainty (EPU) is used to analyse its impact on the overall private consumption as well as the individual impact on its three distinct categories (durable goods, non-durable goods, and services).

A significant contribution is pioneering the examination of economic policy uncertainty indices' significance in predicting private consumption expenditure for USA — a previously unexplored area. Moreover, combined effects of the indicators are also examined.

3. METHODOLOGY

3.1. Data Collection

The overall PCE data for USA as well as its sub-components have been obtained from the Federal Reserve Economic Data (FRED,2023). As per the Bureau of Economic Analysis (BEA), private consumption expenditure is divided into three components: Durable goods, non-durable goods, and services (Bea.gov,2017). The data is available monthly with a lag of one month for the period of 01-January-2004 to 30-April-2023. This time period has been chosen based on the availability of Google trends data.

The data for the Survey-based indicator including Michigan Consumer Sentiment Index (MCSI) is from FRED (2022), and Consumer Confidence Index (CCI) is extracted from the organisation for economic co-operation and development (OECD, 2023). The data is collected monthly with a lag of one month for the period of 01-January-2004 to 30-April-2023. This economic policy uncertainty data has been collected from Policyuncertainty (2012).

The google trends index has been constructed by extracting the monthly data from 01-January-2004 to 30-April-2023. The selection of terms was based on the Vosen & Schmidt (2011) research, by matching the product categories of national income and products account (NIPA) of BEA, by considering 56 google trends categories. Additionally, we have meticulously explored more categories based on the sub-divisions of each private consumption expenditure category as prescribed by BEA and have included 145 google trends categories to this research, as the number of search categories have increased since the time of research of Vosen & Schmidt (2011). The final 145 google trends categories have been classified into durable goods, non-durable goods, and services as per the three categories of BEA. There are 54 categories for durable goods, 41 for non-durable goods and 56 categories for services. Each of the google trends categories represent the overall relative popularity of all Google searches that fall within that category.

Table 1: Matching Google Trends search categories as defined by BEA classification for different components of private consumption expenditure.

Durable goods	Non-durable goods	Services
Art and Craft supplies	Acupuncture and Chinese medicine	Adventure Travel
Arts and entertainment	Alcoholic beverages	Air Travel
Audio equipment	Alternative and natural medicine	Airport Parking and Transportation
Auto financing	Apparel	Apartment and Residential rentals
Auto Insurance	Athletic Apparel	Apparel services

Automotive Industry	Beauty and Fitness	Auto financing
Autos and Vehicles	Beer	Auto insurance
Bicycles accessories	Casual apparel	Bus and rail
Book retailers	Chemicals Industry	Business services
Books and literature	Children's clothing	Car rental and taxi services
Calculators reference tools	Clothing accessories	Carpooling and ridesharing
Camera photo equipment	Coffee and tea	Cleaning supplies and services
CD and Audio shopping	Drugs and medications	College financing
Classic Vehicles	Electricity	Computer Security
Commercial Vehicles	Energy and utilities	Cosmetic Procedure
Communication Equipment	Face and body care	Domestic Services
Computer Electronics	Fast food	E-commerce Services
Computer hardware	Foods and drinks	Education
Computer and Video games	Footwear	Electricity
Cookware and dining ware	Gifts	Entertainment and Media Rentals
Crafts	Grocery and food retailers	Entertainment Media
Custom performance vehicles	Hair Care	Fire and Security Services
DVD and Video shopping	Health	Fitness
Electric plug-in vehicles	Magazines	Food Service
Electronics and Electrical	Makeup and cosmetics	Health Insurance
Entertainment Industry	Men's clothing	Home financing
Eyeglasses and contacts	Newspapers	Home Improvement
Gadgets and portable electronics	Non-alcoholic beverages	Home Insurance
Games	Oil and Gas	Hospitals and Treatment Centres
Gardening and Landscaping	Oral and dental care	Hotels and Accommodations
Home appliances	Pet food and supplies	Housing and Development
Home furnishings	Pharmacy	Insurance
Home Garden	Skin and nail care	Internet and Telecom
Home improvement	Tobacco products	Laundry
Home making and Interior decor	Toys	Legal services
Home storage and shelving	Travel	Mail and package delivery

Hybrid and alternative vehicles	Undergarments	Maritime transport
Internet and Telecom	Vehicle fuels and lubricants	Medical facilities and services
Kitchen and dining	Vision Care	Movies
Luggage and Travel accessories	Wine	Nursing
Major kitchen appliances	Women’s clothing	Parking
Mobile and Wireless accessories		Photo and Video Services
Mobile and Wireless		Rail Transport
Motorcycles		Renewable and Alternative Energy
Music equipment and technology		Restaurants
Off-road vehicles		Retirement and Pension
Office supplies		Security Products and Services
Radio equipment		Social Services
Small kitchen appliances		Spa and Beauty Services
Sporting Goods		Speciality Travel
TV and Video equipment		Travel agencies and Services
Vehicle brands		Travel
Vehicle parts and accessories		Vehicle maintenance
Vehicle shopping		Waste Management
		Water supply and treatment
		Web services

3.2.Data Transformation and Analysis

The extracted datasets were first ensured to be complete without containing any missing values or infinite values, before and after performing any data transformations. The research period selected (2004-2023) carries significant importance due to its inclusion of exceptional economic phases, such as the Great Recession in 2008 and the global COVID-19 pandemic in 2019, followed by an atypical recovery. The dates for all datasets were converted to proper date format. Fig.1 shows the variation in the overall PCE component after seasonal adjustment, which highlights the decline in consumption particularly during the Great Recession and COVID-19, and their associated recoveries. The decompose plot of the overall PCE demonstrates an additive seasonality (Fig.2).

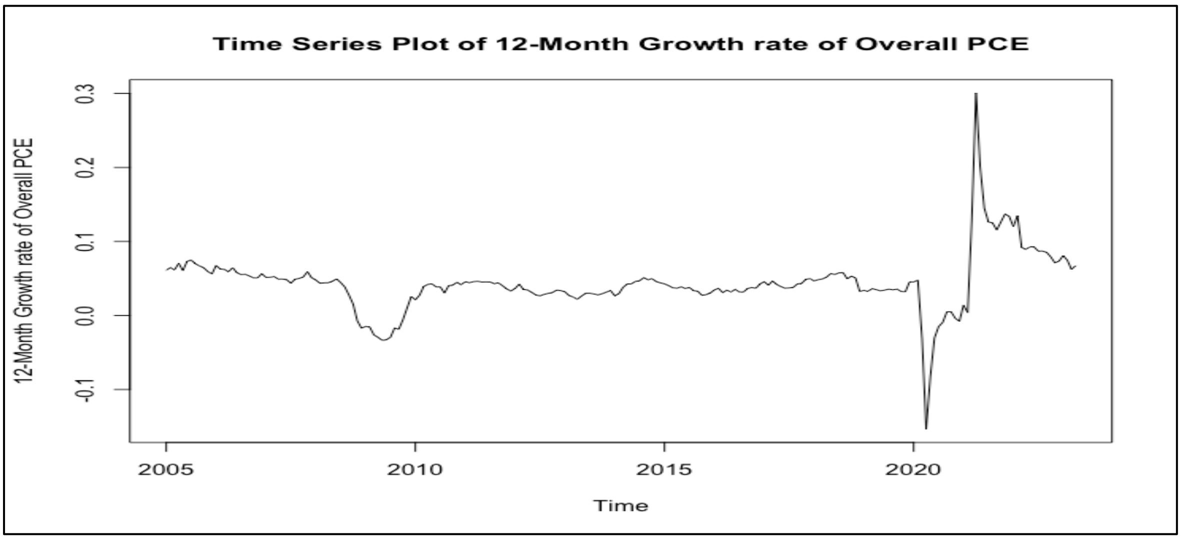


Figure 1: Variation in overall PCE of USA across periods of economic crisis.

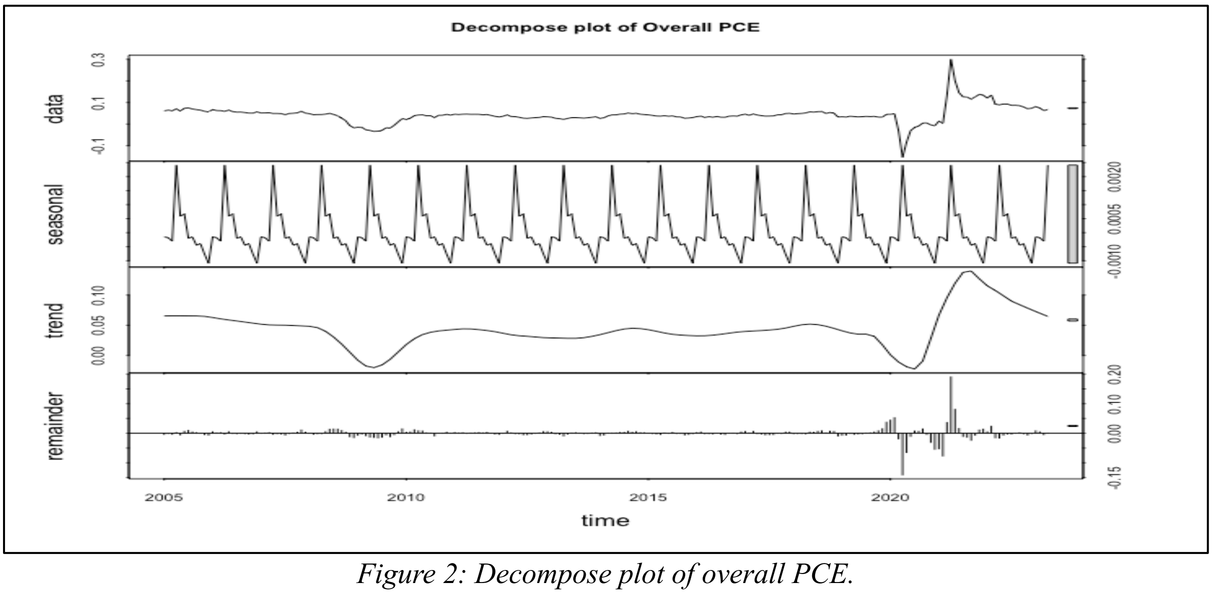


Figure 2: Decompose plot of overall PCE.

The data for Google trends is not seasonally adjusted, hence the 1-month growth rates is converted to 12-month growth rates to account for seasonality. Most of the 145 categories that are in the same component of consumption exhibited high correlation. To address the issue of multicollinearity, Principal Component Analysis (PCA) was employed. The set of factors that collectively account for 90% of the variability in the Google Trends categories associated with each component of consumption was identified, which enables us to capture the most influential patterns and trends while reducing the complexity of the data. After converting the Google Trends data into 12-month growth rates, a similar adjustment was made to the monthly data of overall PCE, PCE categories, MCSI, CCI, and EPU. This seasonal adjustment resulted in a refined timeframe for the data, covering the period from 01-January-2005 to 30-April-2023.

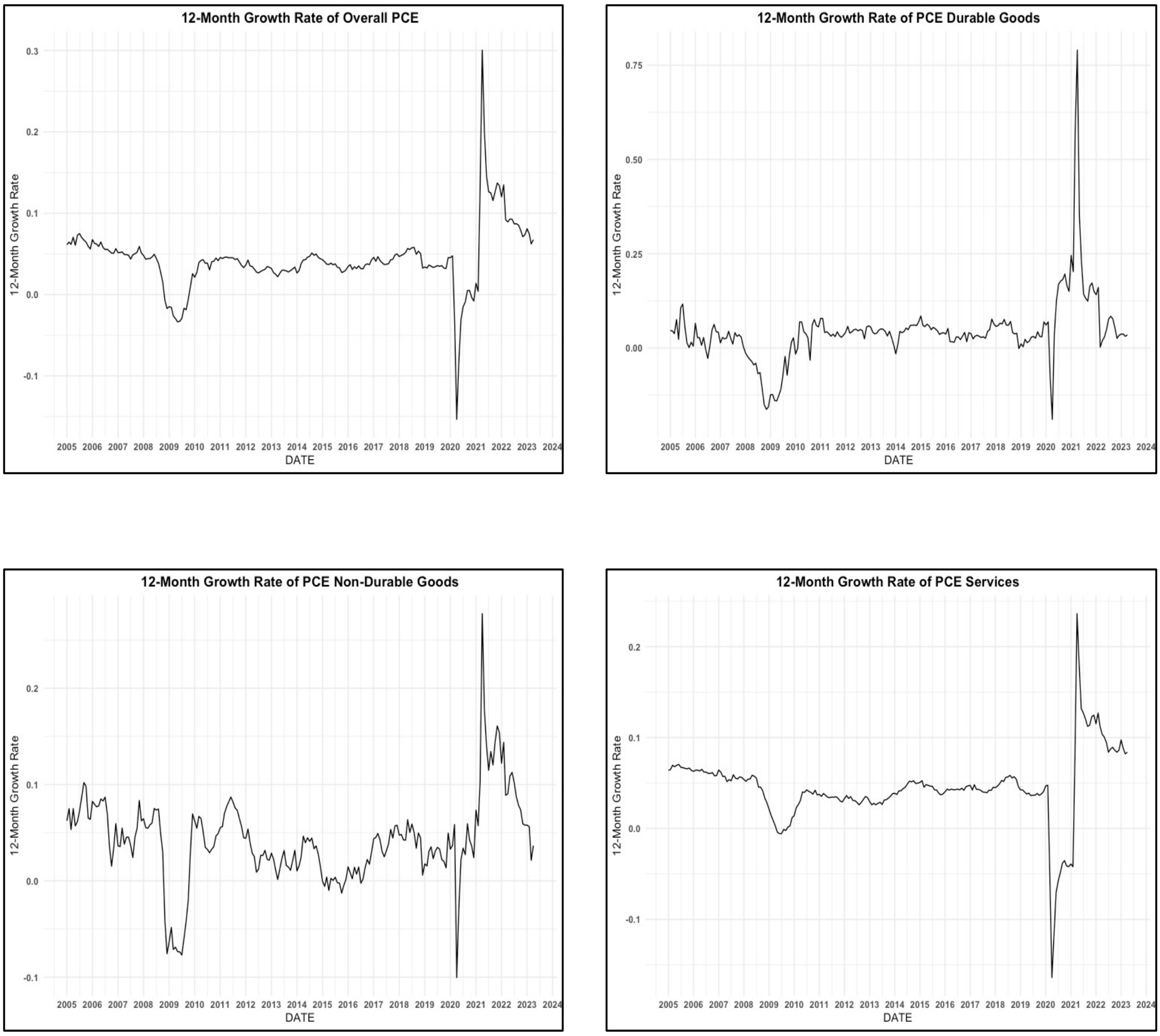


Figure 3: Consumption expenditure trends across various categories displaying decline in consumption during times of economic crisis with high variability in durable and non-durable goods category.

The plot of economic policy uncertainty displays high variability during 2005-2023 suggesting a period of economic instability with significant fluctuations in response to various events policy changes (Fig.4).

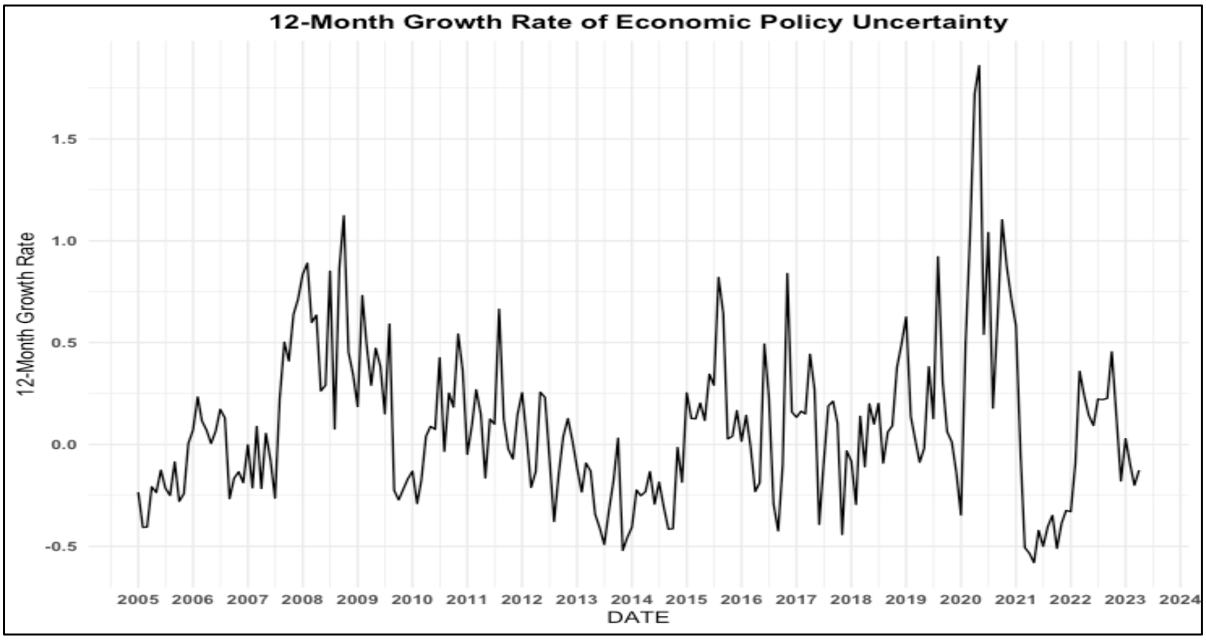


Figure 4: Variation in economic policy uncertainty index across periods of economic crisis.

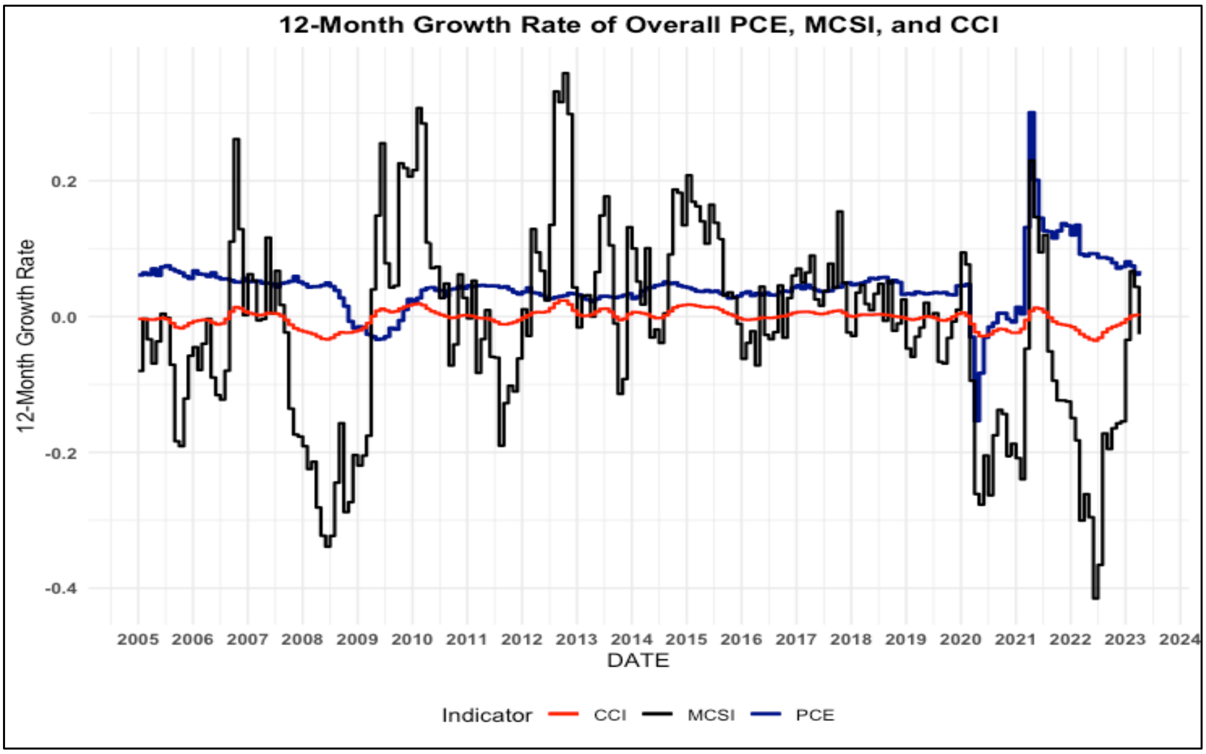


Figure 5: Correlation between overall PCE and survey-based indicators

On analysing the variation of the survey-based indicators (MCSI, CCI) along with the overall PCE demonstrates that the indices follow similar patterns, with the MCSI index showing high variability (Fig.5).

This correlation highlights the power of the indicators for forecasting private consumption expenditure for USA.

3.3. Time Series Forecasting of Private Consumption Expenditure

The entire dataset, covering January,2005 to April,2023, is initially split into two distinct subsets. The first subset pertains to the period prior to the onset of COVID-19, spanning from January,2005 to December,2019. The second subset encompasses the COVID-19 period, ranging from January,2005 to April,2023. This methodology was employed to assess model performance in predicting both the relatively stable pre-COVID data and the tumultuous economic conditions during the COVID-19 crisis.

For the pre-COVID subset, the respective numbers of principal components for each private consumption category are as follows: 19 for Overall PCE, 12 for durable goods (PCEDG), 8 for non-durable goods (PCENDG), and 13 for services (PCES). In the subset covering the COVID-19 period, the numbers of principal components for each private consumption category are: 24 for Overall PCE, 14 for durable goods (PCEDG), 11 for non-durable goods (PCENDG), and 16 for services (PCES).

To assess the extent to which the indicators contribute to predicting variations in consumer spending, we conduct both in-sample and out-of-sample projections. In-sample forecasting involves generating predictions using historical data that the model was trained on, whereas out-of-sample forecasting entails making predictions on data that the model has not been exposed to previously. This approach helps us evaluate the indicators' effectiveness in capturing consumer spending patterns within the dataset and their ability to generalize to new, unseen data.

3.3.1. Forecasting private consumption expenditure using ARIMAX model

The ARIMA (Autoregressive Integrated Moving Average) model constitutes a valuable statistical tool for the analysis of time series data. This model involves three core components: p, d, and q, which respectively symbolize the number of autoregressive (AR) lags, the degree of integration, and the count of moving average (MA) lags. The integration order plays a pivotal role in accommodating non-seasonal variations across periods, thus establishing data stationarity. A time series is deemed stationary when its mean and variance remain consistent over time, and its autocorrelations are contingent solely upon the temporal lag between observations. In this research, the primary objective is to evaluate and compare the predictive accuracy of four distinct models against a baseline Autoregressive Integrated Moving Average (ARIMA) model. This assessment is carried out using historical PCE data.

The baseline model is given by the equation:

$$C_{t+h} = c + \sum_{i=1}^{p} \alpha_i \Delta^d C_{t-i} + \sum_{i=1}^{q} \theta_i \Delta^d \varepsilon_{t-i} + \sum_{i=1}^{P} \phi_i \Delta^D C_{t-si} + \sum_{i=1}^{Q} \eta_i \Delta^D \varepsilon_{t-si} + \varepsilon_t$$

where, C is the monthly 12-month growth rate of the components of private consumption, t is the month at the time of prediction and h is the different forecast horizons. α represents the AR component, θ represents the MA component, ϕ represents the seasonal AR component, η represents the seasonal MA component and ε represents white noise. The parameters p, d, q represents the orders of the autoregressive component, differencing and moving average component, whereas the parameters P, D and Q represents the corresponding order of the seasonal components.

The four models subjected to comparative evaluation encompass the Google Trends model, the University of Michigan's Consumer Sentiment Index (MCSI) model, the Conference Board's Consumer Confidence Index (CCI) model and the Economic Policy Uncertainty (EPU) model. To augment the analytical scope, we also introduce two composite models. The first amalgamates the Google Trends and Economic Policy Uncertainty (EPU) index, while the second combination model incorporates Google Trends, MCSI, CCI, and EPU as combined exogenous regressors. This expansive approach culminates in a total of seven models, including the baseline ARIMA model.

The augmented model is represented as:

$$C_{t+h} = c + \sum_{i=1}^{p} \alpha_i \Delta^d C_{t-i} + \sum_{i=1}^{q} \theta_i \Delta^d \varepsilon_{t-i} + \sum_{i=1}^{P} \phi_i \Delta^D C_{t-si} + \sum_{i=1}^{Q} \eta_i \Delta^D \varepsilon_{t-si} + \sum_{i=1}^{m} \beta_i \mathrm{X}_t + \varepsilon_t$$

where, β is the coefficient of external regressor component and X_t is its value at time t.

The framework begins with a preliminary assessment of stationarity to establish a foundation for subsequent analysis. Subsequently, the order of the ARIMAX model is determined by analysing Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. Ensuring the adequacy of the model's residuals as white noise is a pivotal step in guaranteeing the model's reliability. The accuracy evaluation criterion is based on the Root Mean Square Error (RMSE) for assessing the forecasting precision. The lower the error, the better is the forecasting accuracy of the model.

4. RESULTS AND DISCUSSION

The dataset is divided into two primary subsets: one corresponding to the pre-COVID era spanning from January 2005 to December 2019, and the other encapsulating the COVID era from January 2005 to April 2023.The findings of each segment are explained further. For model assessment, the dataset is partitioned into training and testing sets. During the pre-COVID phase, a division of 75% for training and 25% for testing is implemented, while the including COVID-19 segment adopts an 85-15 split. This augmentation of the split ratio by 10% for the COVID-19 era facilitates a portion of the COVID-19 period for training and the remainder for testing, optimizing forecasting accuracy.

The determination of these optimal split ratios involves several key considerations:

- Given the primary focus on analysing the COVID-19 impact, it is ensured that both the training and testing samples incorporates pertinent COVID-19 data.
- The chosen split ratios are those that exhibit the best fit of the model to actual test data, thereby minimizing the error difference.
- It is ensured that any residual forecast errors display white noise characteristics, signifying robust model performance and comprehensive information capture.

Through meticulous evaluation of these factors, the finalization of distinct split ratios for each subset is established. The analysis encompasses both in-sample and out-of-sample forecasting.

In-sample forecasting accuracy is quantified by measuring the increment in the R-squared value for each model, relative to the baseline R-squared model. The baseline R-squared value is calculated by comparing the variance of the residuals (prediction errors) from the baseline model to the variance of the actual data thereby indicating the proportion of variance in the dependent variables (Overall PCE, PCEDG, PCENDG, PCES) that is explained by the baseline model. The incremental R-squared value holds significance as it provides insight into the extent to which each model improves predictive accuracy beyond the baseline.

Furthermore, the assessment extends to out-of-sample forecasting across different time horizons, quantified using the Root Mean Square Error (RMSE). The Diebold-Mariano p-values are also reported to assess whether there is a statistical difference in forecast accuracy of the augmented models with the baseline (Diebold and Mariano,2002).

4.1.PRE-COVID PERIOD

This sample employs a 75-25 split ratio for training and testing. The training set runs from January,2005 to March,2016. The testing set spans from April,2016 to December,2019.

Table 2: Out-of-sample forecast performance of ARIMAX models in the pre-COVID period as measured by RMSE for various horizons. Diebold Mariano p-values are reported in parenthesis.

OVERALL PCE					
MODEL	**H=3**	**H=6**	**H=9**	**H=12**	**H=15**
BASELINE	**0.00176**	**0.00238**	**0.00248**	**0.00289**	**0.00286**
GOOGLE TRENDS	0.00628 (0.994)	0.00492 (0.992)	0.00441 (0.997)	0.00463 (0.989)	0.00424 (0.988)
MCSI	**0.00179** (0.063)	**0.00237** (0.111)	**0.00247** (0.08)	**0.00287** (0.004)	**0.00285** (0.066)
CCI	**0.00177** (0.134)	**0.00237** (0.133)	**0.00247** (0.04)	**0.00288** (0.076)	**0.00286** (0.189)
EPU	0.0017 (0.985)	0.00250 (0.966)	0.00265 (0.979)	0.00309 (0.972)	0.00306 (0.99)
GOOGLE + EPU	0.00624 (0.997)	0.00481 (0.997)	0.00450 (1.00)	0.00480 (0.996)	0.00437 (0.993)
COMBINATION	0.00727 (0.999)	0.00579 (0.997)	0.00516 (0.999)	0.00510 (0.998)	0.00463 (0.995)
PCEDG					
MODEL	**H=3**	**H=6**	**H=9**	**H=12**	**H=15**
BASELINE	**0.00696**	**0.00601**	**0.01024**	**0.00936**	**0.00841**
GOOGLE TRENDS	0.02958 (0.998)	0.02475 (0.998)	0.02290 (0.999)	0.02047 (0.999)	0.01922 (0.999)
MCSI	**0.0080** (0.901)	**0.00633** (0.819)	**0.01053** (1.00)	**0.00954** (0.785)	**0.00858** (0.815)
CCI	**0.00917** (0.594)	**0.00738** (0.583)	**0.01060** (0.589)	**0.00959** (0.599)	**0.00877** (0.594)
EPU	0.00916 (0.495)	0.00745 (0.493)	0.00948 (0.49)	0.00876 (0.493)	0.00792 (0.496)
GOOGLE + EPU	0.02705 (0.999)	0.02315 (0.999)	0.02116 (1.00)	0.01906 (1.00)	0.01806 (1.00)
COMBINATION	0.04079 (0.979)	0.03350 (0.959)	0.02822 (0.968)	0.02496 (0.952)	0.02340 (0.934)

PCENDG					
MODEL	**H=3**	**H=6**	**H=9**	**H=12**	**H=15**
BASELINE	**0.00459**	**0.00734**	**0.00708**	**0.00654**	**0.00660**
GOOGLE TRENDS	0.00505 (0.991)	0.00821 (1.00)	0.00805 (0.998)	0.00740 (1.00)	0.00746 (1.00)
MCSI	**0.00476** (0.538)	**0.00736** (0.528)	**0.00710** (0.528)	**0.00655** (0.528)	**0.00660** (0.528)
CCI	**0.00555 (0.998)**	**0.00834 (0.998)**	**0.00788 (0.98)**	**0.00720 (0.981)**	**0.00728 (0.963)**
EPU	0.00530 (0.89)	0.00773 (0.878)	0.00699 (0.808)	0.00649 (0.791)	0.00656 (0.791)
GOOGLE + EPU	0.00557 (0.992)	0.00844 (1.00)	0.00756 (0.977)	0.00706 (0.996)	0.00718 (0.997)
COMBINATION	0.00580 (0.993)	0.00856 (1.00)	0.00773 (0.99)	0.00726 (1.00)	0.00733 (1.00)

PCES					
MODEL	**H=3**	**H=6**	**H=9**	**H=12**	**H=15**
BASELINE	**0.00032**	**0.00075**	**0.00141**	**0.00242**	**0.00237**
GOOGLE TRENDS	0.00932 (1.00)	0.01451 (0.997)	0.01241 (0.999)	0.01150 (1.00)	0.01054 (1.00)
MCSI	**0.00037** (0.512)	**0.00075** (0.51)	**0.00139** (0.509)	**0.00240** (0.51)	**0.00235** (0.516)
CCI	**0.00027 (0.657)**	**0.00075 (0.655**	**0.00144 (0.654)**	**0.00244 (0.705)**	**0.00239 (0.717)**
EPU	0.00060 (0.817)	0.00090 (0.875)	0.00156 (0.848)	0.00256 (1.00)	0.00257 (0.829)
GOOGLE + EPU	0.00969 (1.00)	0.01458 (0.999)	0.01268 (1.00)	0.01182 (1.00)	0.01084 (1.00)
COMBINATION	0.00519 (1.00)	0.00420 (1.00)	0.00528 (1.00)	0.00534 (0.999)	0.00482 (0.999)

4.1.1. Overall PCE

The baseline model consistently showcases the strongest performance across all horizons, demonstrating the lowest RMSE values. This suggests its robustness in short-term (H=3,6) as well as long-term forecasting (H=9,12,15), establishing it as the benchmark for accuracy. The survey-based indicators, namely the MCSI and CCI models, closely follow the baseline in terms of predictive power, effectively capturing patterns in the PCE data. Meanwhile, the EPU index exhibits performance comparable to these survey-based indicators, indicating that it contributes to forecasting accuracy in a manner similar to survey-based indicators. In contrast, the Google Trends model and various combination approaches, including Google Trends, display relatively higher RMSE values. This suggests that these models might not effectively capture the intricate nuances of the PCE data, particularly in comparison to the baseline and the survey-based indicator models.

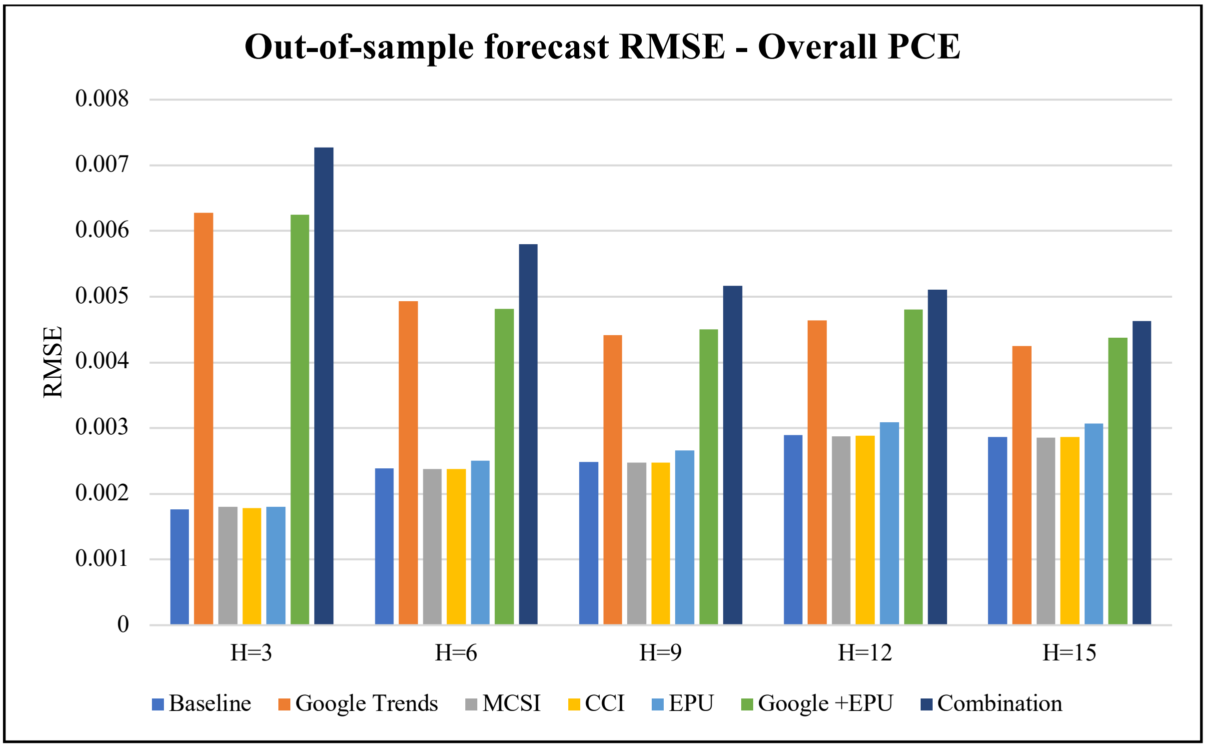


Figure 6: Out-of-sample performance of ARIMAX models in the pre-COVID period for Overall PCE.

4.1.2. PCE Durable Goods (PCEDG)

In the realm of short-term predictions, the Baseline model emerges as the clear leader, displaying consistently low RMSE values across both H=3 and H=6. This model's remarkable ability to capture immediate fluctuations within the PCE durable goods category solidifies its position as the best performer for short-term forecasts. Google Trends model showcases elevated RMSE values compared to survey-based indicators. The MCSI model emerges as the second-best performer for short term predictions.

For long-term predictions (H=9,12,15), the Baseline model maintains its reliability, demonstrating accuracy across extended horizons. However, the EPU model outperforms the baseline in long-term forecasting. Collaborative models like Google + EPU and the Combination model exhibit higher RMSE values in long-term predictions, despite offering a balanced approach.

Both the MCSI and CCI models exhibit consistent proficiency across both short and long-term predictions, presenting relatively low RMSE values. While the MCSI model showcases slightly stronger accuracy in short-term predictions, the differences in RMSE values are relatively modest. Their collective effectiveness underscores the value of incorporating sentiment-driven indices into forecasting models for the durable goods category.

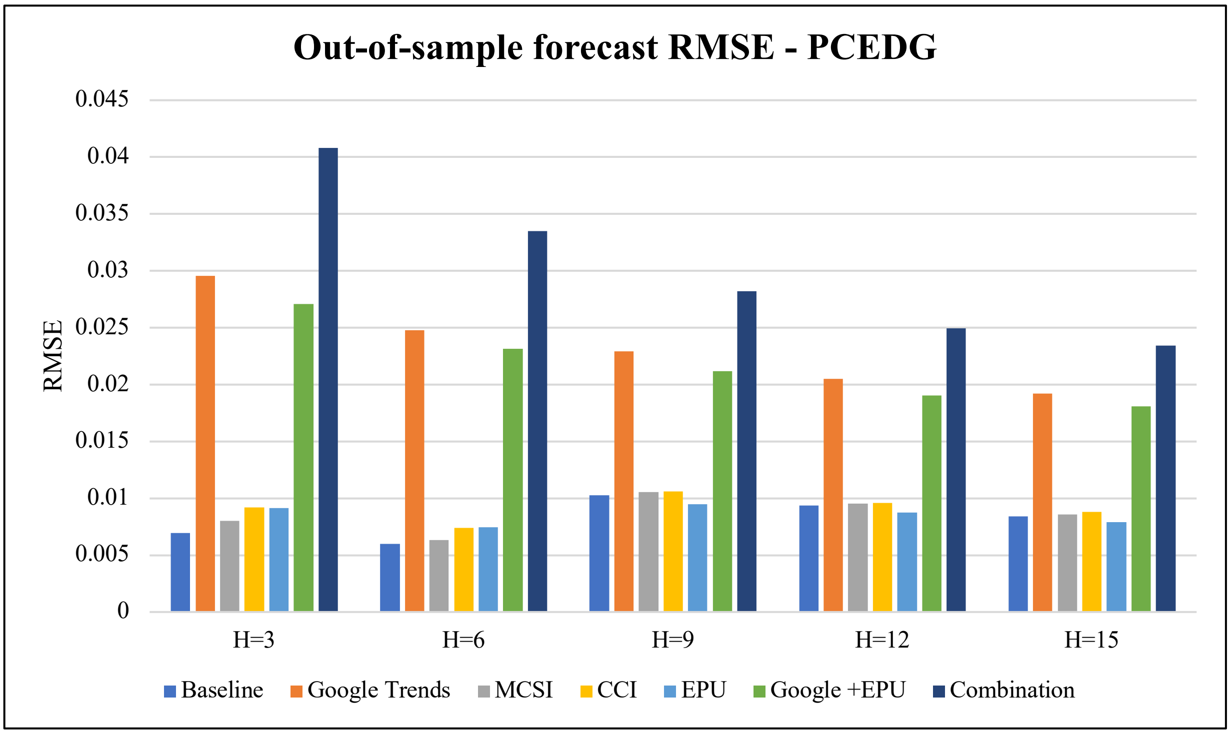


Figure 7: Out-of-sample performance of ARIMAX models in the pre-COVID period for PCE durable goods.

4.1.3. PCE Non-Durable Goods (PCENDG)

When examining the RMSE values for PCE non-durable goods forecasting across different ARIMAX models and time horizons, an interesting observation emerges. The RMSE values for all the models: Baseline, Google Trends, MCSI, CCI, EPU and the combined models are remarkably close. This suggests that the predictive performance of these models is quite comparable, and differences in their forecasting accuracy might fall within a narrow margin. When examining short-term predictions (H=3,6), the Baseline, MCSI and EPU models stand out for their consistent accuracy. These models effectively capture immediate fluctuations in the non-durable goods category, aligning with the inherently dynamic nature of short-term trends.

In the long term (H=9,12,15), a similar trend emerges across models. The Baseline, MCSI and EPU models demonstrate commendable stability, reaffirming their suitability for capturing trends that evolve over extended periods within the non-durable goods landscape. The Google Trends and the combination models presents relatively higher RMSE values. This implies that while it retains competitiveness, there might be some challenges in capturing the intricacies of non-durable goods trends.

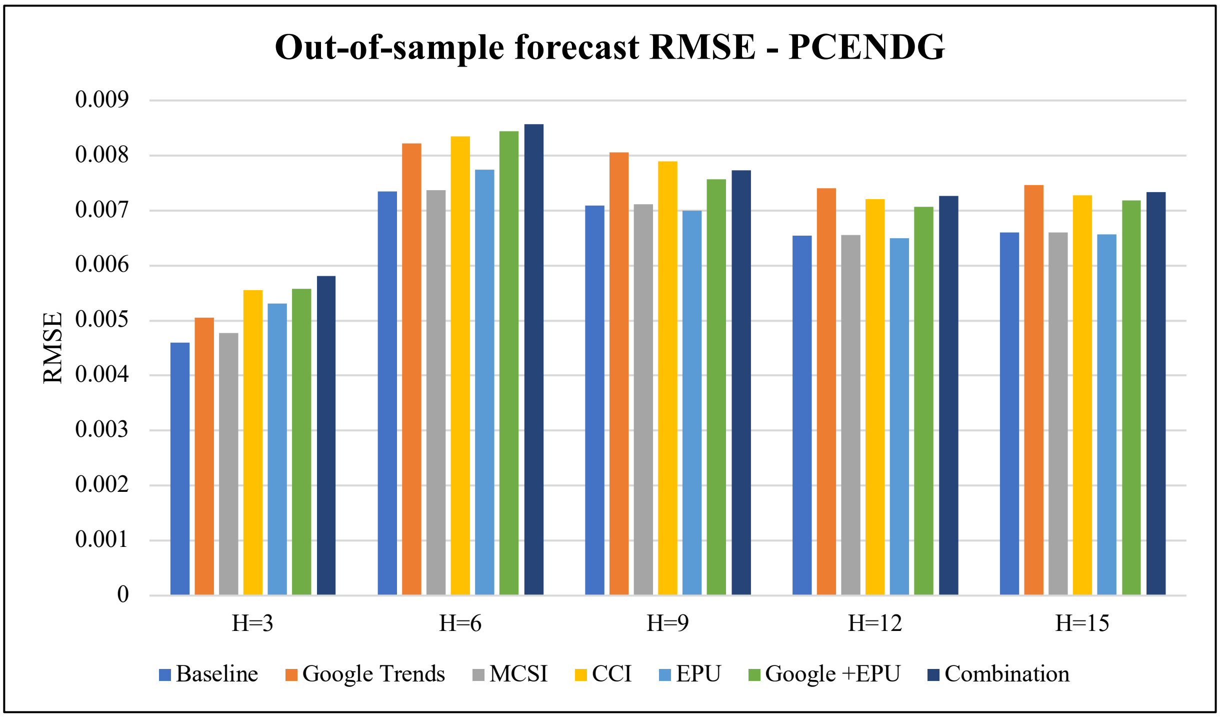


Figure 8: Out-of-sample performance of ARIMAX models in the pre-COVID period for PCE non-durable goods.

4.1.4. PCE Services (PCES)

The baseline model consistently emerges as a strong performer across the entire forecast horizon, suggesting that simple time series methods provide accurate predictions for the PCE services component. This is particularly pronounced in short-term forecasting (H=3,6), where the baseline model displays the lowest RMSE values, showcasing its adaptability to capturing immediate trends.

The EPU index stands out as the best indicator for long-term forecasting of services consumption. The MCSI and CCI models, also demonstrates noteworthy predictive power for long term forecasting. Both MCSI and CCI models consistently yield RMSE values close to the baseline, positioning them as reliable options for both short-term and long-term forecasting. These indices appear to capture sentiment-driven shifts in PCE services consumption effectively. In contrast, models involving Google Trends, Google + EPU, and combined models exhibit higher RMSE values across all horizons. This suggests that these factors might have less impact to accurately predict the PCE services component. While Google Trends appears to have stronger correlation with longer-term trends, the combined models show less consistent performance, indicating the complexity of predicting the services component over time.

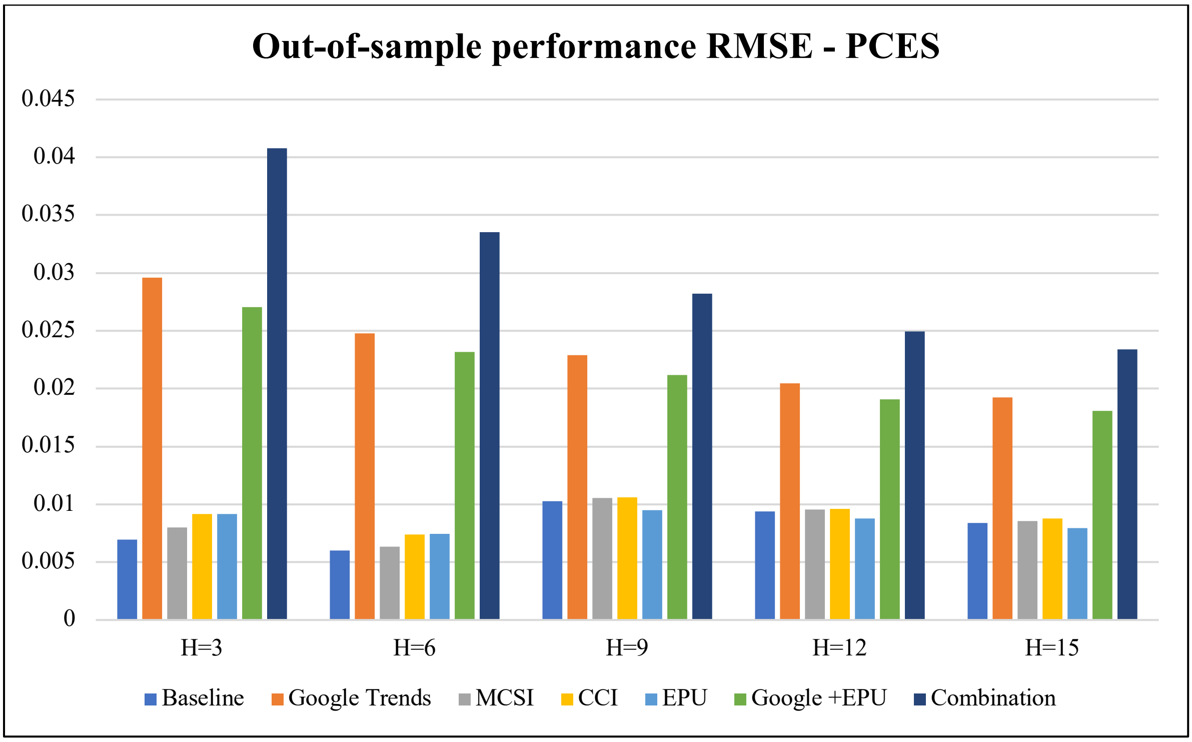


Figure 9: Out-of-sample performance of ARIMAX models in the pre-COVID period for PCE services.

4.1.5. <u>Evaluating In-sample performance</u>

The examination of incremental R^2 values offers valuable insights into the influence of various indicators and their combinations on forecasting consumption patterns. Incorporating Google Trends into the baseline models yields varying outcomes across different consumption components. Notably, it significantly enhances the predictive power for overall private consumption expenditure (PCE-Overall) and durable goods consumption (PCEDG), while its impact on non-durable goods consumption (PCENDG) is negative. Moreover, for services consumption (PCES), Google Trends negatively affects the explanatory power. The addition of the Michigan Consumer Sentiment Index (MCSI) results in modest improvements, suggesting its limited contribution to explaining consumption variation beyond the baseline. Similarly, the Consumer Confidence Index (CCI) produces mixed results: it enhances the model for durable goods consumption, diminishes the explanatory power for non-durable goods consumption, and has minimal impact on services consumption. However, the inclusion of the Economic Policy Uncertainty (EPU) index diminishes the models' explanatory power across the overall consumption and PCENDG. Combining Google Trends with

EPU yields positive impacts on Overall PCE and PCEDG but negatively affects PCENDG and PCES. The Combination model, comprising all indicators, notably augments explanatory power for durable goods consumption, but adversely affects services and non-durable goods consumption.

Table 3: In-sample performance of Baseline model for different components of consumption in the pre-COVID period as measured by R^2 value[2].

BASELINE R^2 VALUE	**OVERALL PCE**	**PCEDG**	**PCENDG**	**PCES**
	96.94	85.57	93.743	97.878

Table 4: In-sample performance of ARIMAX models for different components of consumption in the pre-COVID period as measured by incremental R^2 value with respect to the baseline model.

MODEL	**INCREMENTAL R^2 (MULTIPLIED BY 100)**			
	OVERALL PCE	**PCEDG**	**PCENDG**	**PCES**
GOOGLE TRENDS VS BASELINE	0.175	**1.457**	-1.16	-9.093
MCSI VS BASELINE	0.011	0.152	0.018	0.002
CCI VS BASELINE	-0.003	0.733	-1.321	0.003
EPU VS BASELINE	-0.42	0.096	-0.027	0.095
GOOGLE +EPU VS BASELINE	0.188	**1.54**	-1.043	-8.962
COMBINATION VS BASELINE	**0.31**	**2.85**	-1.031	-9.698

4.1.6. Discussions

During the pre-COVID era, survey-based indicators played a crucial role in determining consumer spending behaviour. The Michigan Consumer Sentiment Index (MCSI) and the Conference Board Consumer Confidence Index (CCI) were two such indices that had consistent predictive power across different consumption categories and forecast horizons. These indices effectively measured consumer sentiment, highlighting that consumer often sought reassurance from their personal financial situations, employment status, and business conditions—captured by survey-based indicators—before making buying choices, especially during periods of economic uncertainty like the Great Recession. The influence of sentiment was reflected on all categories of consumption, particularly when it came to non-durable goods as well as services, where consumers sought the comfort of positive economic outlooks. These findings align with the study of Qiao, McAleer and Wong (2009) who stated the significance of MCSI in predicting consumption movements specifically for non-durable goods. Furthermore, Dees and Soares Brinca (2013) similarly

[2] The baseline R^2 value is calculated by comparing the variance of residuals to the actual data, which indicates the percentage of variation in private consumption expenditure that is explained by the baseline model.

identified that confidence indicators can be good predictors of household consumption in USA particularly during the financial crisis period and they highlighted an “international confidence channel” emphasizing CCI’s role in predicting consumer sentiment across regions. According to our findings, the MCSI index performs slightly better than CCI in terms of short-term forecasting of individual components of PCE, whereas for long term predictions both the indicators are equally significant.

The Baseline model's superiority underlines the fact that it is the foundation for comprehending and forecasting consumer spending patterns. The Baseline's extraordinary capacity to catch rapid changes in consumption patterns confirms its position as the main instrument for short-term and long-term forecasts.

Although Google Trends data has shown less effectiveness in forecasting PCE compared to other indicators, this could be attributed to the selection of search terms that were not prevalent during that period. The study included 145 search terms, but the absence of many terms during the pre-COVID period could have hindered its predictive performance. It’s important to note that previous research (Vosen & Schmidt ,2011; Woo & Oven, 2018) has endorsed the use of relevant search terms during this timeframe, indicating that Google Trends can be a potent predictor. This study doesn’t dismiss the potential of Google Trends but underscores its relative performance when compared to other indicators.

The Economic Policy Uncertainty (EPU) index became an important element affecting consumer spending decisions during periods of economic instability and uncertainty. This index served as a road map for navigating the economic downturn by illuminating how governmental regulations and financial volatility influenced people's decisions. This tendency was vividly demonstrated during the Great Recession, when consumers avidly watched policy changes that might have financial ramifications (Prüser and Schlösser, 2020). This relationship is supported by this study, which emphasises the importance of the EPU index particularly in long-term prediction of consumption expenditure, notably in the consumption of durable items as well as for services, in uncertain times. This complex interaction between behaviour and policy also included the consumption of non-durable items. Consumers skilfully manipulated their spending choices throughout turbulent policy times, firmly basing them in a pragmatic blend of economic stability and policy advances.

4.2.INCLUDING COVID-19 PERIOD

This sample takes the whole dataset split in the ratio 85:15 for training and testing. The training set runs from January,2005 to July,2020. The testing set spans from August,2020 to April,2023.

Table 5: Out-of-sample forecast performance of ARIMAX models in the including COVI-19 period as measured by RMSE for various horizons. Diebold Mariano p-values are reported in parenthesis.

OVERALL PCE					
MODEL	**H=3**	**H=6**	**H=9**	**H=12**	**H=15**
BASELINE	0.01260	0.01150	0.05928	0.05625	0.05056
GOOGLE TRENDS	0.01151 (0.118)	0.01644 (0.09)	**0.01567** **(0.044)**	**0.01933** **(0.025)**	**0.01854** (0.137)
MCSI	0.01126 (0.171)	0.01085 (0.18)	0.05661 (0.174)	0.05324 (0.175)	0.04795 (0.212)
CCI	**0.00718** (0.209)	**0.00871** (0.231)	0.05426 (0.243)	0.05184 (0.25)	0.04684 (0.239)
EPU	0.01277 (0.228)	0.01155 (0.24)	0.06004 (0.239)	0.05635 (0.362)	0.05059 (0.064)
GOOGLE + EPU	0.01127 (0.118)	0.0162 (0.09)	**0.01581** **(0.044)**	**0.01955** **(0.025)**	**0.01877** (0.137)
COMBINATION	0.01116 (0.118)	0.01610 (0.091)	**0.01579** **(0.046)**	**0.01937** **(0.027)**	**0.01857** (0.136)
PCEDG					
MODEL	**H=3**	**H=6**	**H=9**	**H=12**	**H=15**
BASELINE	0.00657	0.03311	0.13878	0.14253	0.13515
GOOGLE TRENDS	0.05770 (0.116)	0.06069 (0.187)	**0.11249** (0.135)	**0.12429** (0.098)	**0.11294** **(0.004)**
MCSI	0.00631 (0.223)	0.03273 (0.203)	0.13236 (0.168)	0.13214 (0.019)	0.12975 (0.203)
CCI	**0.00614** (0.443)	**0.03120** (0.437)	0.12892 (0.41)	0.13463 (0.325)	0.13281 (0.354)
EPU	0.00880 (0.766)	0.03352 (0.706)	0.13736 (0.784)	0.13985 (0.894)	0.13708 (0.768)
GOOGLE + EPU	0.05698 (0.121)	0.06007 (0.193)	**0.11214** (0.146)	**0.12390** (0.112)	**0.11255** **(0.002)**
COMBINATION	0.05978	0.06217	**0.10955**	**0.12374**	**0.11275**

	(0.138)	(0.216)	(0.163)	(0.143)	**(0.002)**
PCENDG					
MODEL	**H=3**	**H=6**	**H=9**	**H=12**	**H=15**
BASELINE	0.01454	0.01854	0.05088	0.04799	0.04312
GOOGLE TRENDS	0.01647 (0.141)	0.02427 (0.092)	**0.03792** (0.126)	**0.03624** (0.126)	**0.03420** (0.121)
MCSI	**0.01385** (0.405)	0.01811 (0.397)	0.04920 (0.399)	0.04404 (0.383)	0.040441 (0.422)
CCI	**0.01408** (0.508)	**0.01826** (0.509)	0.04977 (0.509)	0.04472 (0.52)	0.04108 (0.507)
EPU	0.01570 (0.634)	0.01860 (0.667)	0.05221 (0.624)	0.04656 (0.845)	0.04250 (0.607)
GOOGLE + EPU	0.01645 (0.14)	0.02431 (0.092)	**0.03783** (0.125)	**0.03616** (0.125)	**0.03413** (0.12)
COMBINATION	0.01647 (0.134)	0.02436 (0.083)	**0.03878** (0.113)	**0.03648** (0.113)	**0.03434** (0.118)
PCES					
MODEL	**H=3**	**H=6**	**H=9**	**H=12**	**H=15**
BASELINE	0.00717	0.00625	0.04569	0.04407	0.04260
GOOGLE TRENDS	**0.00616** (0.049)	0.01036 (0.056)	**0.0142** (0.009)	**0.01379** (0.002)	**0.01253** (0.077)
MCSI	0.00985 (0.336)	0.00834 (0.263)	0.05039 (0.308)	0.04838 (0.351)	0.04354 (0.328)
CCI	**0.00628** (0.218)	**0.0057** (0.271)	0.03877 (0.227)	0.03756 (0.235)	0.03668 (0.284)
EPU	0.00717 (0.191)	0.00625 (0.3)	0.04569 (0.009)	0.04407 (0.002)	0.04260 (0.077)
GOOGLE +EPU	0.00611 (0.049)	0.01032 (0.056)	**0.01422** (0.009)	**0.01381** (0.002)	**0.01254** (0.077)
COMBINATION	0.00614 (0.049)	0.01034 (0.056)	**0.01423** (0.01)	**0.01379** (0.002)	**0.01248** (0.078)

4.2.1. Overall PCE

In terms of short-term and long-term forecasting, the RMSE values indicate the models' performance across various horizons. For short-term forecasting (H=3, 6), the CCI model stands out with the lowest RMSE, showcasing its accuracy in predicting immediate changes. Following closely is the Google Trends and MCSI model for H=3. Interestingly, for H=6 the EPU model performs slightly better than Google trends. This indicates that the survey-based indicators as well as the Google trends and EPU indicators can be a reliable source for short term predictions of the overall consumption patterns.

For long-term forecasting (H=9,12,15) the Combination model emerges as the most accurate choice also supported by the Diebold-Mariano values. This suggests that the combined approach performs well in predicting the extended trends that unfold over a longer time span.

Comparing the different models, the CCI and Google Trends models display robust performance, with the former excelling in short-term forecasting and the latter being competitive across various horizons. Notably, the MCSI and EPU models present similar performance, demonstrating consistent RMSE values across different horizons.

When comparing the survey-based indicators, MCSI and CCI, the data indicates that the CCI hold more predictive power in forecasting overall PCE.

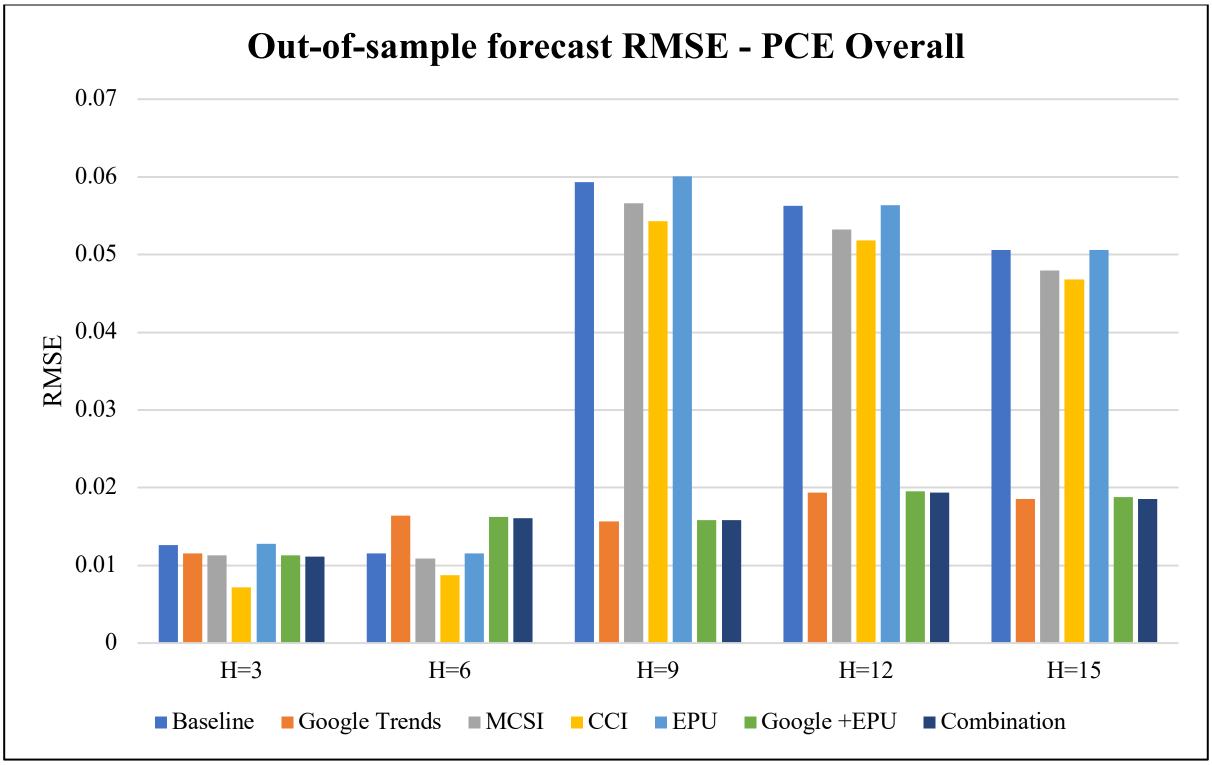


Figure 10: Out-of-sample performance of ARIMAX models in the including COVID-19 period for overall PCE.

4.2.2. <u>PCE Durable Goods (PCEDG)</u>

In terms of short-term forecasting, particularly at Horizons H=3 and H=6, the CCI model takes the lead with the lowest RMSE values. This highlights its adeptness in capturing immediate fluctuations in PCE durable goods, closely followed by the competitive MCSI model and the EPU model.

However, for long-term forecasting, the picture shifts. The Combination model emerges as the most accurate choice indicating its strength in predicting extended trends in PCE durable goods consumption. Google Trends also performs admirably in long-term forecasting, surpassing both MCSI and CCI in this aspect. This suggests that Google Trends is indeed effective in capturing longer-term consumer spending shifts.

Meanwhile, the EPU model exhibits results that might not be as strong when compared to some of the other models. However, its significance becomes more apparent when considering the collaborative power of combining it with Google Trends.

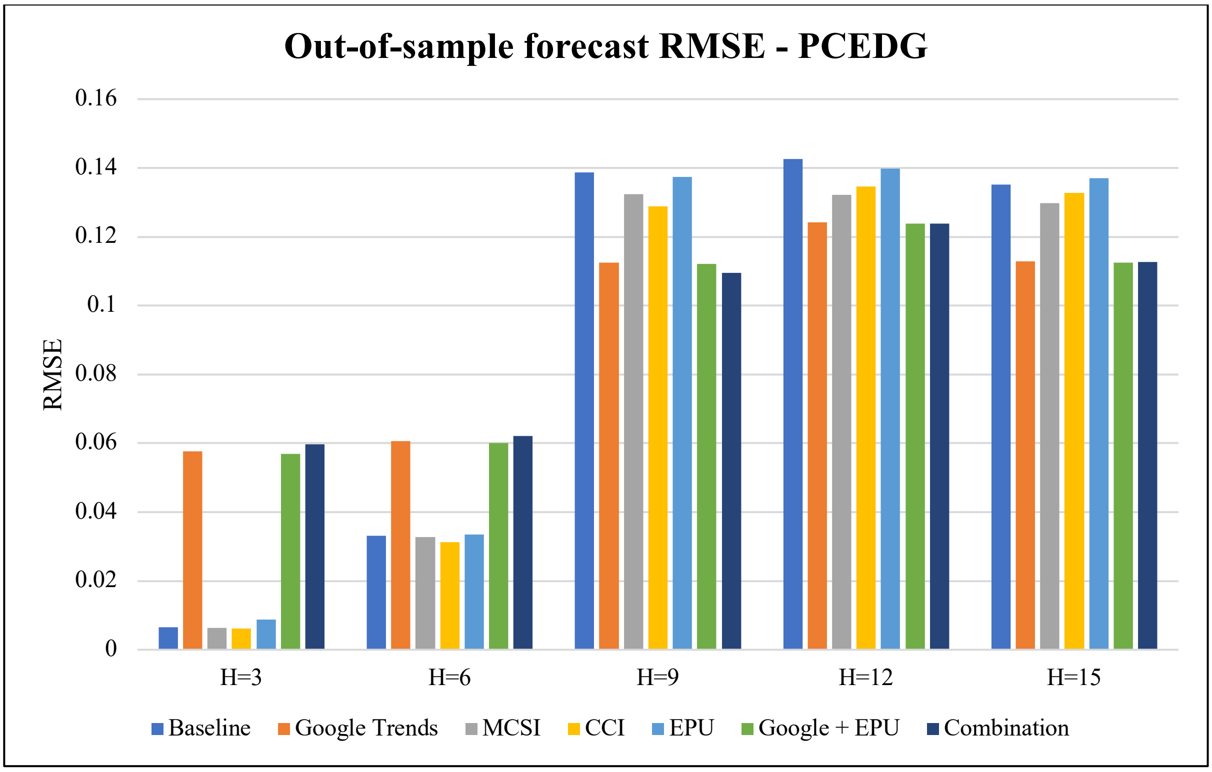


Figure 11: Out-of-sample performance of ARIMAX models in the including COVID-19 period for PCE durable goods.

4.2.3. PCE Non-Durable Goods (PCENDG)

The RMSE values for PCE non-durable goods forecasting demonstrate consistent and competitive performance across various ARIMAX models. The proximity of RMSE values suggests that the models are generally effective in predicting this category of consumption. Upon closer inspection, the CCI model emerges as a consistent performer. With RMSE values that range closely around 0.014, it showcases its competence in capturing immediate fluctuations in PCE non-durable goods consumption. Following suit, the MCSI and Baseline models maintain similar accuracy with RMSE values in the same range as the CCI model for short-term predictions. Google Trends, however, demonstrates slightly higher RMSE values, suggesting that its effectiveness might vary in the short term for this category of consumption.

Transitioning to long-term forecasting, the results display an intriguing pattern. The Combination model surfaces as a strong candidate, performing consistently well with RMSE values around 0.034 to 0.038. This suggests that the combined approach, leveraging multiple indicators, excels in capturing extended trends in PCE non-durable goods consumption. Google Trends notably improves its performance in long-term forecasting. The EPU model, when combined with Google Trends as the Google + EPU model, contributes to enhanced predictive capabilities. While its individual performance might not be the most accurate, its synergy with online search insights creates a more comprehensive forecasting strategy, particularly suited for PCE non-durable goods.

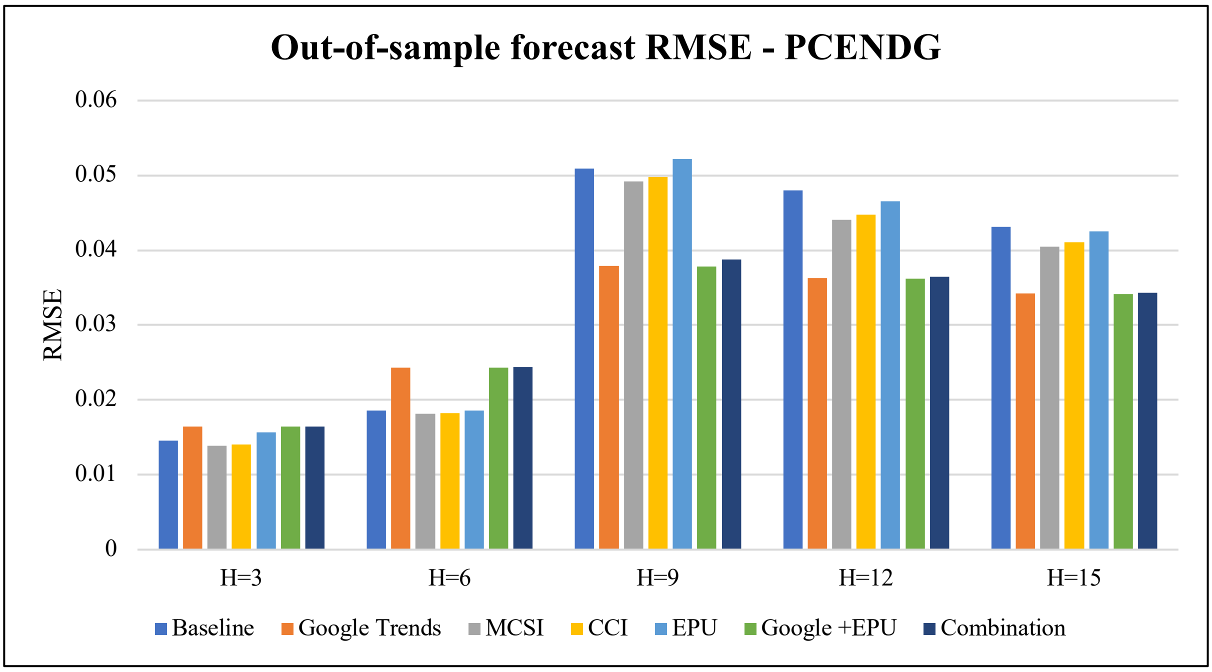


Figure 12: Out-of-sample performance of ARIMAX models in the including COVID-19 period for PCE non-durable goods.

4.2.4. <u>PCE Services (PCES)</u>

The CCI model continues to shine as the best performer, consistently demonstrating remarkable accuracy with the lowest RMSE values. Its ability to swiftly capture immediate fluctuations in the PCE services category solidifies its position as a reliable choice for short-term forecasting. Interestingly the Google Trends index similar performance with CCI in this category of consumption for H=3. In tandem with the CCI model's excellence, the MCSI and EPU index performs consistently well for short term forecasting. The Baseline model remains a reliable benchmark, providing a stable foundation for predictions across short term periods.

The Combination model emerges as the best performer, showcasing unparalleled accuracy with consistently low RMSE values across long time periods. Moreover, the Google + EPU model secures the second-best position for long-term predictions. Reflecting on the EPU model, while it maintains steady performance, its accuracy relative to other models, especially in long-term forecasts, falls slightly behind. However, a transformative effect comes to light when it collaborates with Google Trends, as evidenced by the Google + EPU model's refined and comprehensive short-term prediction capabilities.

Comparing the strengths of MCSI and CCI, both models exhibit notable accuracy, particularly in the short term. However, MCSI introduces a touch of variability in long-term forecasts, contrasting with the sustained excellence that CCI demonstrates over extended horizons.

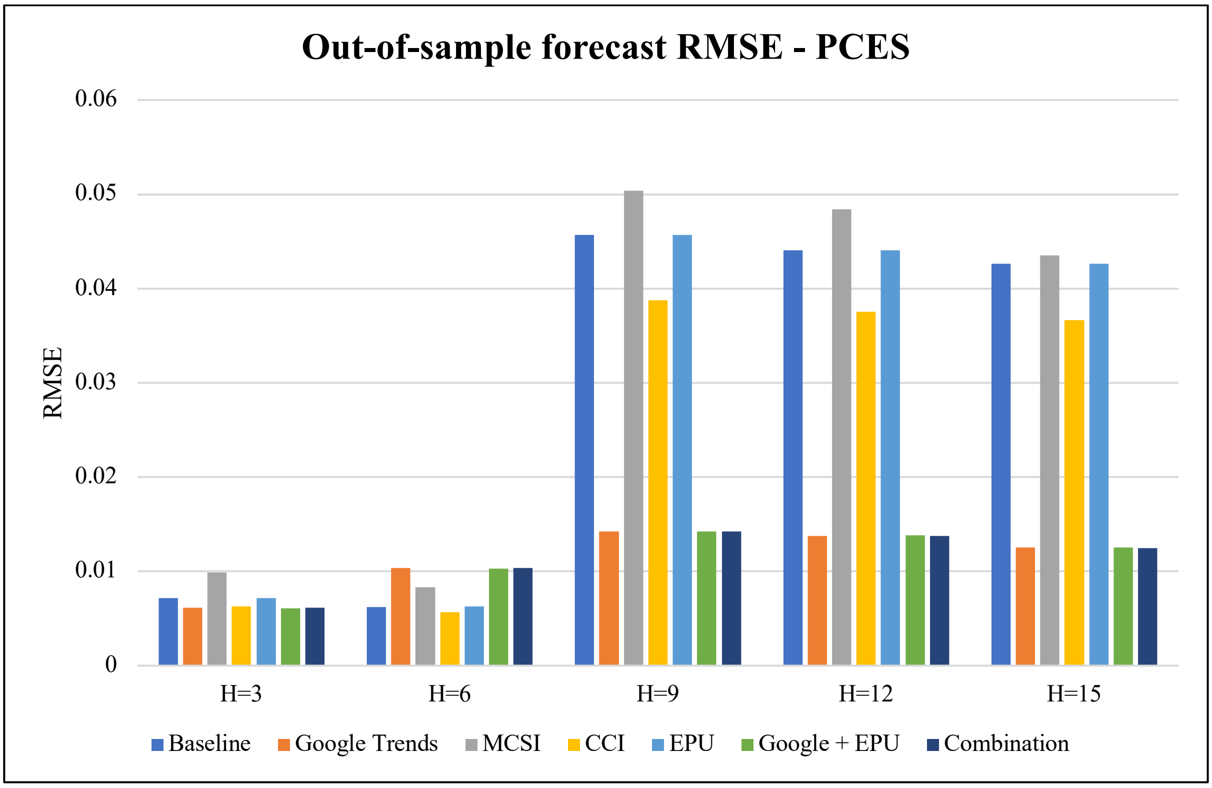


Figure 13: Out-of-sample performance of ARIMAX models in the including COVID-19 period for PCE services.

4.2.5. Evaluating In-sample performance

The baseline R^2 values, derived from ARIMA models with historical PCE values of corresponding components, serve as the starting point for comparison. Predictive power is significantly increased when Google Trends data is incorporated into the baseline models. Google Trends makes a considerable contribution to all consumption components (Overall PCE, PCEDG, PCENDG, and PCES), with particularly notable gains seen for PCE-Overall and PCES. Positive incremental R^2 values for Overall PCE, PCEDG, and PCENDG are recorded by the MCSI, which increases predictive power, albeit to a lower extent than Google Trends. It's interesting to note that while MCSI has a favourable impact on the consumption of durable and non-durable products, it has very little of a negative effect on the consumption of services.

Across all consumption categories, the Consumer Confidence Index produces positive incremental R2 values, demonstrating a general improvement in predicting accuracy. This shows that variations in spending habits, particularly when it comes to the consumption of durable and services goods, are explained in part by consumer attitude, as measured by the CCI. The majority of consumption components show positive incremental R^2 values for the EPU index, which is linked to economic instability, suggesting that it makes a minimal contribution to forecasting accuracy. However, the effect on service consumption (PCES) is nearly negligible.

Predictive power is significantly improved when Google Trends and the EPU index are combined, especially for Overall PCE, PCEDG and PCENDG. This combination shows significant gains, because of the complementarity between Google Trends, consumer sentiment and the EPU index's policy-related uncertainty. The pattern seen with Google Trends and the joint Google Trends + EPU model is mirrored by the combination model, which includes all indicators, and boosts the explanatory value for all consumption components uniformly.

Table 6: In-sample performance of Baseline model for different components of consumption in the including COVID-19 period as measured by R^2 value[3].

BASELINE R^2 VALUE	**OVERALL PCE**	**PCEDG**	**PCENDG**	**PCES**
	79.6	74.022	79.285	86.984

Table 7: In-sample performance of ARIMAX models for different components of consumption in the including COVID-19 period as measured by incremental R^2 value with respect to the baseline model.

MODEL	**INCREMENTAL R^2 (MULTIPLIED BY 100)**			
	OVERALL PCE	**PCEDG**	**PCENDG**	**PCES**
GOOGLE TRENDS VS BASELINE	**16.232**	**5.952**	**4.592**	**9.565**
MCSI VS BASELINE	0.431	1.036	1.13	-0.79
CCI VS BASELINE	1.138	1.253	0.975	1.152
EPU VS BASELINE	0.139	0.521	1.102	2.35093e-06
GOOGLE +EPU VS BASELINE	**16.249**	**6.038**	**4.603**	0.9654955
COMBINATION VS BASELINE	**16.273**	**6.553**	**4.645**	**9.595**

[3] The baseline R^2 value is calculated by comparing the variance of residuals to the actual data, which indicates the percentage of variation in private consumption expenditure that is explained by the baseline model.

4.2.6. Discussions

The dynamics of consumer spending predictions have been impacted in the period that includes the COVID-19 timeframe by both the ongoing global health problem and the interaction of numerous economic factors. The efficiency of different forecasting models and indicators has changed because of the addition of real-world events and economic disturbances.

The importance of Google Trends as a prediction tool is highlighted in this period. It is noteworthy that Google Trends shows promise as a leading indicator for both short-term and long-term forecasts in predicting all components of consumption particularly for overall PCE and PCEDG. The Diebold-Mariano p-values are also significant. This insight can be applied beyond the pandemic, as real-world events and economic disturbances continue to shape consumer spending patterns. This is supported by the McKinsey & Company survey conducted by Alldredge *et al.* (2022) which highlights the impact of online shopping behaviour during COVID-19 in USA revealing that consumers have not only embraced but continued to use e-commerce channels, with enthusiasm in categories such as food, non-food products and home improvement, even as they resumed out-of-home activities. It's evident that Google Trends, by tracking online search patterns and consumer preferences, could effectively capture the evolving consumer behaviours and preferences across all categories of consumption in the era of online shopping.

Both the Consumer Confidence Index (CCI) and the Michigan Consumer Sentiment Index (MCSI) consistently project accuracy, especially in the short term. The COVID-19 outbreak, among other real-world events, has enhanced CCI's ability to identify sudden shifts in consumer spending patterns, notably in categories like PCE Services. Our data (Fig.3) reveals declines in consumer confidence and sentiment during the COVID-19 period from 2020-2021 and a recent significant decline in 2022. The COVID-19 decline can be attributed to virus resurgences, lockdown fears, and job security concerns (Anne Smith,2021). The 2022 decline was linked to reduced comfort in major and household purchases, along with worries about inflation, bills, and taxes (Ipsos,2022). Alongside these challenges, the consistency of the CCI model in projecting accuracy suggests its relevance in predicting sudden changes in consumer spending, particularly in the short term. This decline reflects consumers' concerns about their economic prospects and demonstrates how CCI model can serve as a valuable indicator for immediate spending decisions.

The EPU index's overall effectiveness continues to be noteworthy, especially when combined with Google Trends. This implies that policy-related uncertainty and internet search findings can work well together to provide a more comprehensive understanding of consumption expenditure behaviour. This was particularly well-suited for overall PCE and PCEDG, where consumers' spending decisions were influenced by a combination of economic instability and policy-related uncertainty. Moreover, our results reveal a subtle but significant variation in short-term predictions, particularly at horizon H=6, across all consumption categories. In this context, the EPU index exhibits a distinct advantage over the MCSI and Google Trends. Upon closer analysis, this specific horizon (H=6) coincides with the onset of the third wave of the COVID-19 pandemic in the United States.

The significance of cooperation between indicators is yet another important finding. Both the joint Google Trends + EPU model and the Combination model have demonstrated improved prediction skills across all categories of consumption. When there is major economic uncertainty, the interaction of various variables seems to offer a more complete picture of consumer spending behaviour.

5. CONCLUSION

This research delved into the complex world of private consumption expenditure (PCE) forecasting for USA by analysing the impacts of various economic indicators, spanning both the pre-COVID (2005-2019) and COVID-19 (2005-2023) periods. The thorough analysis exposes a multi-faceted narrative of the private consumption behaviour for the overall PCE as well as its individual components, and the profound effects of unrivalled economic crises.

i. Impact of Economic Indicators on ARIMAX Model for PCE Forecasting:

   Prior to COVID, survey-based indicators showcased consistent predictive power for short-term forecasting which included the Consumer Confidence Index (CCI) and Michigan Consumer Sentiment Index (MCSI), while Google Trends was less promising for detecting consumption expenditure trends. The CCI retained its short-term prediction prowess during the COVID-19 era, while Google Trends and Combination models emerged as the most accurate option for long-term forecasting, demonstrating a greater relevance of collaborative forecasting techniques in uncertain times.

ii. Evolution of Consumption Patterns and COVID-19's Impact:

   Consumption patterns in the United States evolved significantly from 2005 to 2023, shaped by pivotal events like the Great Recession and COVID-19. Pre-COVID (2005-2019), durable and non-durable goods consumption remained stable, supported by the various models including the baseline model. Services consumption also showed stability, aligning with consistent consumer demand. The baseline model consistently supported this stability in overall PCE and its individual components during this period. However, the COVID-19 period disrupted these patterns, as numerous models indicated reduced durable goods consumption due to lockdowns. Non-durable goods consumption declined initially, reflecting shifting priorities. Services consumption underwent significant changes during COVID-19, with multiple models suggesting notable shifts with respect to the baseline model, especially during lockdowns. Consumers increasingly turned to online resources like Google Trends to inform their service-related purchase decisions. In comparison to the pre-COVID era, when offline indications had a greater influence on predicting consumer behaviour, this digital transformation represented a substantial shift.

iii. Significance of EPU in Predicting PCE and its Interactions:

   The EPU index significantly influenced private consumption expenditure in both pre-COVID and COVID-19 eras, particularly for long-term forecasts. During the pre-pandemic period, it played a substantial role in shaping consumer decisions, especially for long-term predictions for durable goods and services. In the COVID-19 era, the EPU index continued to impact consumer sentiment and expenditure, with enhanced predictive capabilities when combined with Google Trends data. Survey-based indicators, MCSI and CCI, also interacted with the EPU index to shape consumer sentiment and spending patterns, with significant impacts depending on consumption categories and forecast timelines.

iv. Combined Effects of Economic Indicators on PCE Forecasting:

Intricate differences between the pre-COVID and COVID-19 periods were seen in the combined influence of economic indices. Amidst the COVID-19 period, the Combined Model demonstrated remarkable enhancements in forecasting private consumption expenditure (PCE), particularly excelling in long-term predictions. This phenomenon underscored the intricate interplay among economic indices, consumer sentiment, and online search behaviour. In contrast, in the pre-COVID era, the combined model was less suitable as opposed to survey-based indicators for both short term as well as long term prediction of consumption expenditure.

This study's implications have significant impact for stakeholders facing situations of economic uncertainty like the COVID-19 pandemic. Indicators of economic policy uncertainty can be used by policymakers to anticipate changes in private consumption expenditure and make well-informed decisions. Businesses are advised to focus on short-term strategies, making modifications based on survey-based indicators like the Consumer Confidence Index. With its real-time insights and ability to make both short- and long-term predictions, Google Trends stands out as essential. These insights essentially act as a practical road map for decision-makers to quickly adjust plans, aligning with customer preferences and economic realities, thereby successfully navigating risks.

5.1. Limitations

This study's limitation is the variable effectiveness of Google Trends, which is dependent on the search terms used. The choice of search keywords, which can affect the precision of predictions, is intricately tied to Google Trends' effectiveness as a forecasting tool. Google Trends data is frequently updated, producing dynamic and shifting scores that could affect the performance of the model. Furthermore, even while Google Trends offers insightful statistics on users' online search habits, there is some subjectivity introduced by the dataset's lack of definition for category selection. This restriction emphasises the need of selecting search keywords carefully and underlines the dynamic nature of data on internet search behaviour.

5.2. Future work

Future research could investigate ways to improve private consumption expenditure (PCE) forecast by looking more closely at the contributions of other economic variables. The baseline model can be expanded to include historical PCE data as well as macroeconomic variables like stock prices, disposable income, and treasury bills. Given that the ARIMA model is univariate, attempts can be made to accommodate such extensions by investigating multivariate models like Vector Autoregression (VAR). The inclusion of machine learning techniques is a tempting field for investigation as the forecasting landscape evolves. Future research may also focus on evaluating other Google search categories and assess how they function in times of economic crisis.

6. REFERENCES

Abdulazeez, S.A., 2021. A comparative study between ARIMA and ARIMAX in forecasting gross domestic product (GDP) in Nigeria. *KASU Journal of Mathematical Science*, *2*(2), pp.30-42

Al-Thaqeb, S.A., Algharabali, B.G. and Alabdulghafour, K.T. (2020). The pandemic and economic policy uncertainty. International Journal of Finance & Economics. doi:https://doi.org/10.1002/ijfe.2298.

Alldredge, K. *et al.* (2022) *US consumer spending: 2022 survey results McKinsey*, *www.mckinsey.com*. [Online]. [Accessed 27 Aug 2023]. Available at: https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/how-us-consumers-are-feeling-shopping-and-spending-and-what-it-means-for-companies

Anne Smith, K. 2021. *Consumer Confidence Drops As Covid-19 Infections Rise Across U.S.*, *Forbes Advisor* . Edited by K. Bailie. [Online]. [Accessed 27 August 2023].

Available at: https://www.forbes.com/advisor/personal-finance/consumer-confidence-falls-delta-variant-rises/

Baker, S.R., Bloom, N. and Davis, S.J. (2013). *Measuring Economic Policy Uncertainty*. [Online] papers.ssrn.com. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2198490.

Baker, S.R., Farrokhnia, R.A., Meyer, S., Pagel, M. and Yannelis, C., 2020. How does household spending respond to an epidemic? Consumption during the 2020 COVID-19 pandemic. The Review of Asset Pricing Studies, 10(4), pp.834-862.

Balcilar, M., Gupta, R. and Segnon, M. (2016). The Role of Economic Policy Uncertainty in Predicting U.S. Recessions: A Mixed-frequency Markov-switching Vector Autoregressive Approach. *Economics: The Open-Access, Open-Assessment E-Journal*. doi:https://doi.org/10.5018/economics-ejournal.ja.2016-27.

Bea.gov. (2017). *NIPA Handbook: Concepts and Methods of the U.S. National Income and Product Accounts | U.S. Bureau of Economic Analysis (BEA)*. [Online] [Accessed 20 June 2023]. Available at: https://www.bea.gov/resources/methodologies/nipa-handbook.

Binder, C.C., 2016. Economic policy uncertainty and household inflation uncertainty. *The BE Journal of Macroeconomics*, *17*(2), p.20160048.

Bloom, N., 2009. The impact of uncertainty shocks. *econometrica*, *77*(3), pp.623-685.

Bonsu,C.O.andMuzindutsi,P.F.,2017.Macroeconomic determinants of household consumption expenditure in Ghana: A multivariate cointegration approach. International Journal of Economics and Financial Issues, 7(4),pp.737-745.

Bram, J. and Ludvigson, S.C., 1998. Does consumer confidence forecast household expenditure? A sentiment index horse race. Economic Policy Review, 4(2).

Caggiano, G., Castelnuovo, E. and Figueres, J.M., 2017. Economic policy uncertainty and unemployment in the United States: A nonlinear approach. *Economics Letters*, *151*, pp.31-34.

Carnot, N., Koen, V., Tissot, B., 2005. Economic Forecasting. Springer.

Carnot, Nicolas, Koen, V., Tissot, B., 2005. Time Series Methods, in: Carnot, N., Koen, V., Tissot, B. (Eds.), Economic Forecasting. Palgrave Macmillan UK, London, pp. 85–102. https://doi.org/10.1057/9780230005815_4

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

Carroll,C.D.,1994.How does future income affect current consumption?.The Quarterly Journal of Economics, 109(1), pp.111-147.

Cavallo, A., 2020. Inflation with Covid consumption baskets (No. w27352). National Bureau of Economic Research. doi:https://doi.org/10.3386/w27352

CEICdata.com (2018). United States Private Consumption: % of GDP. [Online]. [Accessed 3 July 2023]. Ceicdata.com. Available at: https://www.ceicdata.com/en/indicator/united-states/private-consumption--of-nominal-gdp.

Choi, H. and Varian, H., 2009. Predicting initial claims for unemployment benefits. GoogleInc, 1(2009), pp.1-5.

Choi, H. and Varian, H., 2012. Predicting the present with Google Trends. Economic record, 88, pp.2-9.

Choi, H., Varian, H., 2012. Predicting the Present with Google Trends. Econ. Rec. 88, 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Chron.com. (2020). The Importance of Consumer Spending. [Online]. [Accessed 12 July 2023].
Available at: https://smallbusiness.chron.com/importance-consumer-spending-3882.html.

Curtin,R.T.,1982.Indicators of consumer behavior:TheUniversityof Michigan surveys of consumers. Public Opinion Quarterly, 46(3), pp.340-352.

Dees, S. and Soares Brinca, P. 2013. Consumer confidence as a predictor of consumption spending: Evidence for the United States and the Euro area. *International Economics*, 134, pp.1–14. doi:https://doi.org/10.1016/j.inteco.2013.05.001.

Della Penna, N. and Huang, H., 2009. Constructing consumer sentiment index for US using internet search patterns. Department of Economics, WP, 26.

Diebold, F.X. and Mariano, R.S., 2002. Comparing predictive accuracy. *Journal of Business & economic statistics*, *20*(1), pp.134-144.

Eppright, D.R., Arguea, N.M. and Huth, W.L. (1998). Aggregate consumer expectation indexes as indicators of future consumer expenditures. Journal of Economic Psychology, 19(2), pp.215–235. doi:https://doi.org/10.1016/s0167-4870(98)00005-1

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

Fajar, M., Prasetyo, O.R., Nonalisa, S. and Wahyudi, W., 2020. Forecasting unemployment rate in the time of COVID-19 pandemic using Google trends data (case of Indonesia).

Fasulo, A., Guandalini, A. and Terribili, M.D., 2017. Google Trends for nowcasting quarterly household consumption expenditure. Rivista Italiana di Economia Demografia e Statistica, 71(4), p.5.

FRED (2022). *FRED*. Stlouisfed.org. [Online]. [Accessed 3 July 2023]. Available at: https://fred.stlouisfed.org.

Fulton, C. and Hubrich, K. (2021). Forecasting US Inflation in Real Time. Finance and Economics Discussion Series.2021(014), pp.1–32. [Online]. [Accessed 20 July 2023]. doi:https://doi.org/10.17016/feds.2021.014.

Gil, M., Pérez, J.J., Sanchez Fuentes, A.J. and Urtasun, A., 2018. Nowcasting private consumption: traditional indicators, uncertainty measures, credit cards and some internet data.

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, 1012–1014. https://doi.org/10.1038/nature07634

Howrey, E.P., 2001. The predictive power of the index of consumer sentiment. Brookings papers on economic activity, 2001(1), pp.175-207.

Hyndman, R.J. and Athanasopoulos, G. (2018). *Forecasting : Principles and Practice*. 2nd ed. Heathmont, Otexts.

Ingle, C., Bakliwal, D., Jain, J., Singh, P., Kale, P. and Chhajed, V. (2021). Demand Forecasting: Literature Review On Various Methodologies. IEEE Xplore. [Online]. [Accessed 20 July 2023]. doi:https://doi.org/10.1109/ICCCNT51525.2021.9580139.

Ipsos. 2022. Overall U.S. consumer sentiment falls to second lowest point of 2022. Ipsos. https://www.ipsos.com/en-us/news-polls/overall-us-consumer-sentiment-falls-second-lowest-point-2022

Islam, T.U. and Mumtaz, M.N., 2016. Consumer Confidence Index and Economic Growth: An Empirical Analysis of EU Countries. EuroEconomica, 35(2).

Kim, J., 2016. Consumer confidence and economic activity: What causes what?. Korea and the World Economy, 17(2), pp.183-212.

Krollner, B., Vanstone, B. and Finnie, G. (2010). Financial time series forecasting with machine learning techniques: European Symposium on Artificial Neural Networks. Proceedings of the 18th European Symposium on Artificial Neural Networks (ESANN 2010). pp.25–30. [Online]. [Accessed 28 July

2023]. Available at: https://research.bond.edu.au/en/publications/financial-time-series-forecasting-with-machine-learning-technique.

Laaksoharju, M. (2011). Correlation and causality between consumer sentiment and stock markets: Evidence from SP500. Master's Thesis, Vaasa University.

Law,T.J.(2019).19 Powerful Ecommerce StatisticsThat Will Guide Your Strategy in 2020. www.oberlo.com. [Online]. [Accessed 20 March 2023]. Available at: https://www.oberlo.com/blog/ecommerce-statistics.

Leduc, S. and Liu, Z., 2016. Uncertainty shocks are aggregate demand shocks. *Journal of Monetary Economics*, *82*, pp.20-35.

Ludvigson,S.C.,2004.Consumerconfidenceandconsumer spending. Journal of Economic perspectives, 18(2), pp.29-50.

Marketsandmarkets.com. (2016). Big Data Market By Software, Services & Deployment Mode. MarketsandMarkets. [Online] Available at: https://www.marketsandmarkets.com/Market-Reports/big-data-market-1068.html.

Mazurek, J. and Mielcová, E., 2017. Is consumer confidence index a suitable predictor of future economic growth? An evidence from the USA. Economics and Management.

Mishkin, F.S., Hall, R., Shoven, J., Juster, T. and Lovell, M. (1978). Consumer Sentiment and Spending on Durable Goods. Brookings Papers on Economic Activity, 1978(1), p.217. doi:https://doi.org/10.2307/2534366.

Mowen, J.C., Young, C.E. and Silpakit, P., 1985. A Comparison of the University of Michigan and Conference Board Indices of Consumer Economic Attitudes. ACR North American Advances.

Munyaradzi Ganyaupfu, E. (2020). Health services costs, household income and health expenditure in South Africa. *Journal of Applied Economics and Business*, 8(3), pp.20–30

OECD.2023. *Leading indicators - Consumer confidence index (CCI) - OECD Data*. OECD. [Online]. [Accessed 5 July 2023]. Available at: https://data.oecd.org/leadind/consumer-confidence-index-cci.htm.

Peter, Ď. and Silvia, P., 2012, September. ARIMA vs. ARIMAX–which approach is better to analyze and forecast macroeconomic time series. In *Proceedings of 30th international conference mathematical methods in economics* (Vol. 2, pp. 136-140).

Phan, D.H.B., Iyke, B.N., Sharma, S.S. and Affandi, Y. (2020). Economic policy uncertainty and financial stability–Is there a relation? *Economic Modelling*. [Online] doi:https://doi.org/10.1016/j.econmod.2020.02.042.

Piger, J.M. 2003. Consumer Confidence Surveys: Do They Boost Forecasters' Confidence?. Stlouisfed.org. [Online]. [Accessed 30 June 2023]. Available at: https://www.stlouisfed.org/publications/regional-economist/april-2003/consumer-confidence-surveys-do-they-boost-forecasters-confidence

Policyuncertainty.com. 2012. *Economic Policy Uncertainty Index*. [Online]. [Accessed 4 July 2023]. Available at: https://www.policyuncertainty.com/us_monthly.html.

Prüser, J. and Schlösser, A. 2020. On the Time-Varying Effects of Economic Policy Uncertainty on the US Economy. *Oxford Bulletin of Economics and Statistics*, 82(5), pp.1217–1237. doi:https://doi.org/10.1111/obes.12380.

Qiao, Z., McAleer, M. and Wong, W.-K.2009. Linear and nonlinear causality between changes in consumption and consumer attitudes. *Economics Letters*, 102(3), pp.161–164. doi:https://doi.org/10.1016/j.econlet.2008.12.010.

Serafini, G., Yi, P., Zhang, Q., Brambilla, M., Wang, J., Hu, Y., Li, B., 2020. Sentiment-Driven Price Prediction of the Bitcoin based on Statistical and Deep Learning Approaches, in: 2020 International Joint Conference on Neural Networks (IJCNN). Presented at the 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. https://doi.org/10.1109/IJCNN48605.2020.9206704

Shim, S., Eastlick, M.A., Lotz, S.L. and Warrington, P., 2001. An online prepurchase intentions model: the role of intention to search: best overall paper award—The Sixth Triennial AMS/ACRA Retailing Conference, 2000. Journal of retailing, 77(3), pp.397-416.

To, P.L., Liao, C. and Lin, T.H., 2007. Shopping motivations on Internet: A study based on utilitarian and hedonic value. Technovation, 27(12), pp.774-787.

Vosen,S.andSchmidt,T.,2011.Forecastingprivateconsumption:survey- based indicators vs. Google trends. Journal of forecasting, 30(6), pp.565-578.

Woo, J. and Owen, A.L. 2018. Forecasting private consumption with Google Trends data. Journal of Forecasting, 38(2), pp.81–91.

Zhang, C., Tian, Y.-X., Fan, Z.-P., Liu, Y. and Fan, L.-W. (2019). Product sales forecasting using macroeconomic indicators and online reviews: a method combining prospect theory and sentiment analysis. Soft Computing. doi:https://doi.org/10.1007/s00500-018-03742-1.

7. APPENDICES

7.1.Sample R code

```
#-------------------------------------------------------------------------------------------------------------------
#--------------------------------------DATA PRE-PROCESSING------------------------------------------------
#-------------------------------------------------------------------------------------------------------------------

#---------------------Google Trends ---------------------------

# Get a list of all CSV files in the working directory
file_list <- list.files(pattern = "*.csv", full.names = TRUE)

# Create an empty list to store the loaded data frames
data_frames <- list()

# Load each CSV file into a data frame and add it to the list
for (file in file_list) {
  # Read the CSV file skipping the first two rows
  data <- read.csv(file, skip = 2)

  # Rename the "Month" column to "DATE"
  names(data)[names(data) == "Month"] <- "DATE"

  # Convert the "DATE" column to date format ("%Y-%m" format)
  data$DATE <- as.Date(paste(data$DATE, "-01", sep = ""), format = "%Y-%m-%d")

  # Rename the "Geo: United States" column to "Frequency"
  names(data)[names(data) == "Geo..United.States"] <- "Frequency"

  # Calculate the 12-month growth rate
  data$`Month_12_Growth_Rate` <- 0
  for (i in 13:nrow(data)) {
    data$`Month_12_Growth_Rate`[i] <- (data$Frequency[i] - data$Frequency[i - 12]) / data$Frequency[i -
12]
  }

  # Remove the first year of data (first 12 rows)
  data <- data[-c(1:12), ]

  # Add the data frame to the list
  data_frames[[file]] <- data
}

# Access the loaded data frames
# Example: printing the structure of each data frame
for (file in file_list) {
```

```
  print(paste("File:", file))
  print(str(data_frames[[file]]))
}

# Get the number of files in the dataframe
file_count <- length(file_list)
cat("Number of files in the dataframe:", file_count, "\n")

# Get the count of data in each file
for (file in file_list) {
  data <- data_frames[[file]]
  data_count <- nrow(data)
  cat("Data count in", file, ":", data_count, "\n")
}

# Check for NA values and infinity values in each file
for (file in file_list) {
  data <- data_frames[[file]]

  # Check for NA values
  if (any(is.na(data))) {
    cat("File", file, "contains NA values.\n")
  }

  # Check for infinity values
  if (any(is.infinite(data$Month_12_Growth_Rate))) {
    cat("File", file, "contains infinity values.\n")
  }
}

#-----------------Saving the updated files -----------------
# Create a new directory for the output files
output_folder <- "/Users/name/Desktop/dissertation\ datasets/updated_non_durable"

# Save the updated files to the new folder
for (file in file_list) {
  # Extract the file name without extension
  file_name <- tools::file_path_sans_ext(basename(file))

  # Create the output file path
  output_file <- file.path(output_folder, paste0(file_name, "_updated.csv"))

  # Save the updated data frame as CSV
  write.csv(data_frames[[file]], file = output_file, row.names = FALSE)

  cat("Saved updated file:", output_file, "\n")
}

#---------------------------------MCSI--------------------------
# Read the CSV file
```

```
mcsi <- read.csv("MCSI.csv")

summary(mcsi)
head(mcsi)
# Convert "DATE" column to proper date format
mcsi$DATE <- as.Date(mcsi$DATE, format = "%d/%m/%Y")
head(mcsi)
# Calculate the 12-month growth rate
mcsi$`Month_12_Growth_Rate` <- 0
for (i in 13:nrow(mcsi)) {
  mcsi$`Month_12_Growth_Rate`[i] <- (mcsi$UMCSENT[i] - mcsi$UMCSENT[i - 12]) / mcsi$UMCSENT[i -
12]
}
# Save the updated data as "mcsi_updated.csv" excluding the first 12 rows
write.csv(mcsi[13:nrow(mcsi), ], file = "mcsi_updated.csv", row.names = FALSE)

#-----------------------------PCE DATA ------------------------
pce <- read.csv("PCE.csv")
summary(pce)
head(pce)
# Convert "DATE" column to proper date format
pce$DATE <- as.Date(pce$DATE, format = "%d/%m/%Y")
pce$`Month_12_Growth_Rate` <- 0
for (i in 13:nrow(pce)) {
  pce$`Month_12_Growth_Rate`[i] <- (pce$PCE[i] - pce$PCE[i - 12]) / pce$PCE[i - 12]
}
pce
# Save the updated data as "pce_updated.csv" excluding the first 12 rows
write.csv(pce[13:nrow(pce), ], file = "pce_updated.csv", row.names = FALSE)


#------------------------------CCI DATA ------------------
cci <- read.csv("CCI.csv")
summary(cci)
# Remove all columns except "TIME" and "Value"
cci <- cci[, c("TIME", "Value")]
# Rename the "TIME" column to "DATE"
colnames(cci)[colnames(cci) == "TIME"] <- "DATE"
class(cci$DATE)
head(cci)
# Append "-01" to all the values in the "DATE" column
cci$DATE <- paste(cci$DATE, "01", sep = "-")
cci$DATE <- as.Date(cci$DATE, format = "%Y-%m-%d")
head(cci)
cci$`Month_12_Growth_Rate` <- 0
for (i in 13:nrow(cci)) {
  cci$`Month_12_Growth_Rate`[i] <- (cci$Value[i] - cci$Value[i - 12]) / cci$Value[i - 12]
}
cci
# Save the updated data as "cci_updated.csv" excluding the first 12 rows
```

```
write.csv(cci[13:nrow(cci), ], file = "cci_updated.csv", row.names = FALSE)


#----------------------------US Economic Policy Uncertainty -----------
uncertainty <- read.csv("US_Policy_Uncertainty_Data_updated.csv")
head(uncertainty)
# Combine "Year" and "Month" columns to form a date column
uncertainty$DATE <- as.Date(paste(uncertainty$Year, uncertainty$Month, "01", sep = "-"))
# Remove the Year and Month columns
uncertainty <- uncertainty[, c("DATE", "Three_Component_Index")]
class(uncertainty$DATE)
uncertainty$`Month_12_Growth_Rate` <- 0
for (i in 13:nrow(uncertainty)) {
  uncertainty$`Month_12_Growth_Rate`[i] <- (uncertainty$Three_Component_Index[i] -
uncertainty$Three_Component_Index[i - 12]) / uncertainty$Three_Component_Index[i - 12]
}
uncertainty
# Save the updated data as "uncertainty_updated.csv" excluding the first 12 rows
write.csv(uncertainty[13:nrow(uncertainty), ], file = "uncertainty_updated.csv", row.names = FALSE)

#---------------------------Overall Google Trends ------------------------------
# Set the working directory to the folder containing .csv files
# Replace the path below with the actual path on your system
setwd("/Users/name/Desktop/dissertation/updated_google_trends")

# Get the list of all .csv files in the folder
csv_files <- list.files(pattern = "\\.csv$")

# Initialize an empty list to store the data frames
data_list <- list()

# Loop through each .csv file, read it, and store the selected columns in the list
for (file in csv_files) {
  data <- read.csv(file, header = TRUE, stringsAsFactors = FALSE)
  selected_data <- data[, c("DATE", "Month_12_Growth_Rate")]
  data_list[[file]] <- selected_data
}
# Loop through each data frame in data_list and convert "DATE" to date format
for (i in seq_along(data_list)) {
  data_list[[i]]$DATE <- as.Date(data_list[[i]]$DATE)
  data_list[[i]]$Month_12_Growth_Rate <- as.numeric(data_list[[i]]$Month_12_Growth_Rate)
}

# Check for NA values in each individual data frame in data_list
for (i in seq_along(data_list)) {
  cat("\nFile:", csv_files[i], "\n")
  print(anyNA(data_list[[i]]))
  # Check for infinite values in the "Month_12_Growth_Rate" column of the current data frame
  print(any(is.infinite(data_list[[i]]$Month_12_Growth_Rate)))
}
```

```
# Create a master date vector containing all unique "DATE" values from all data frames
master_dates <- unique(unlist(lapply(data_list, function(df) df$DATE)))

# Loop through each data frame and identify missing dates
missing_dates_by_file <- list()

for (i in seq_along(data_list)) {
  file_dates <- data_list[[i]]$DATE
  missing_dates <- setdiff(master_dates, file_dates)
  if (length(missing_dates) > 0) {
    missing_dates_by_file[[csv_files[i]]] <- missing_dates
  }
}

# Print the missing dates by file
if (length(missing_dates_by_file) > 0) {
  cat("\nMissing dates by file:\n")
  for (file in names(missing_dates_by_file)) {
    cat("File:", file, "\n")
    cat("Missing Dates:", missing_dates_by_file[[file]], "\n\n")
  }
} else {
  cat("\nNo missing dates found in any file.\n")
}


# Define a custom function to perform a full join on the DATE column
custom_join <- function(df1, df2) {
  # Use full_join() function from the dplyr package to perform the join
  # The by argument specifies the column on which the join will be performed (in this case, "DATE")
  inner_join(df1, df2, by = "DATE")
}

library(dplyr)
library(purrr)
# Perform the join operation using reduce from dplyr
combined_data <- reduce(data_list, custom_join)
# Sort the data frame based on the "DATE" column
combined_data <- combined_data[order(combined_data$DATE), ]
# Filter the data for the specified date range
end_date <- as.Date("2019-12-01")
binded_dataframe <- combined_data[combined_data$DATE <= end_date, ]
binded_dataframe <- binded_dataframe[,c(-1)]

#binded_dataframe <- combined_data[,c(-1)] #including covid
anyNA(binded_dataframe)

#Performing PCA analysis
pca <- prcomp(binded_dataframe, center=TRUE, scale = TRUE)
```

```
summary(pca) #choose components that explain more than 90% of the variance - cumulative proportion
#plotting the elbow curve
plot(pca, type="line", main = "PCA of Overall Google Trends")

pca$x
pca_google_trends_pre <- pca$x
pca_google_trends_pre <- pca_google_trends_pre[,c(1:15)]
pca_google_trends_pre <- as.data.frame(pca_google_trends_pre)

# Save the updated data as "pca_google_trends.csv"
write.csv(pca_google_trends_pre, file = "pca_google_trends.csv", row.names = FALSE)

#------------------------------------------------------------------------------------------------------------------
#--------------------------------------DATA ANALYSIS -----------------------------------------------------------
#------------------------------------------------------------------------------------------------------------------

pce <- read.csv("pce_updated.csv")
summary(pce)
pce$DATE <- as.Date(pce$DATE)

head(pce)
library(ggplot2)
library(scales)
library(forecast)

summary(pce$PCE)

ggplot(pce, aes(x = DATE, y = PCE)) +
  geom_line(color = "black") +
  scale_x_date(date_labels = "%Y", date_breaks = "1 year") +  # Format date labels to show only years
  labs(title = "Overall Private Consumption Expenditure (PCE) Over Time",
       x = "Date",
       y = "PCE") +
  theme_minimal()


ts_data <- ts(pce$Month_12_Growth_Rate, start = c(2005, 1), end = c(2023, 4), frequency = 12)
# Create a time series plot with customizations
plot(ts_data, ylab = "12-Month Growth rate of Overall PCE", main = "Time Series Plot of 12-Month
Growth rate of Overall PCE", col = "black")


#Decompose stl
decompose_pce <- stl(ts_data,s.window="periodic")
plot(decompose_pce,main="Decompose plot of Overall PCE")


#install.packages("ggplot2")
library(ggplot2)
head(pce)
```

```
class(pce$DATE)
ggplot(pce, aes(x = DATE, y = Month_12_Growth_Rate)) +
  geom_line(color = "black") +
  ylab("12-Month Growth Rate") +
  ggtitle("Date vs 12-Month Growth Rate of Overall PCE") +
  scale_x_date(date_labels = "%Y", date_breaks = "1 year") +
  theme_minimal() +
  theme(axis.text = element_text(face = "bold"))

mcsi <- read.csv("mcsi_updated.csv")
cci <- read.csv("cci_updated.csv")
mcsi$DATE <- as.Date(mcsi$DATE)
cci$DATE <- as.Date(cci$DATE)
head(mcsi)
head(cci)

# Calculate correlation matrix
cor_matrix <- cor(cbind(pce$PCE, mcsi$UMCSENT, cci$Value))

# Print the correlation matrix
print(cor_matrix)

library(corrgram)

# Create a corrgram for the correlation matrix
corrgram(cor_matrix)


# Create a combined data frame
combined_data <- data.frame(DATE = pce$DATE,
                            PCE = pce$Month_12_Growth_Rate,
                            MCSI = mcsi$Month_12_Growth_Rate,
                            CCI = cci$Month_12_Growth_Rate)

# Melt the data frame into long format
combined_data_long <- reshape2::melt(combined_data, id.vars = "DATE", variable.name = "Indicator",
value.name = "Growth_Rate")

library(reshape2)

ggplot(combined_data_long, aes(x = DATE, y = Growth_Rate, color = Indicator)) +
  geom_step(data = combined_data_long[combined_data_long$Indicator == "PCE", ], size = 1) +
  geom_step(data = combined_data_long[combined_data_long$Indicator != "PCE", ],
            aes(group = Indicator), size = 0.8) +
  scale_color_manual(values = c(PCE = "darkblue", MCSI = "black", CCI = "red")) +
  ylab("12-Month Growth Rate") +
  ggtitle("12-Month Growth Rate of Overall PCE, MCSI, and CCI") +
  scale_x_date(date_labels = "%Y", date_breaks = "1 year") +
  theme_minimal() +
  theme(legend.position = "bottom")+
```

```
  theme(axis.text = element_text(face = "bold"),
        plot.title = element_text(hjust = 0.5, face = "bold"))





#------------------------------------------------------------------------------------------------------------------------
#--------------------------DATA MODELLING: INCLUDING COVID-19 PERIOD---------------------------
#------------------------------------------------------------------------------------------------------------------------

# Install and load necessary packages
#install.packages("forecast")
library(forecast)
#install.packages("TSstudio")
library(TSstudio)
library(tseries)

#Reading the PCE data
pce <- read.csv("pces_updated.csv")
head(pce)
# Convert DATE column to Date format
pce$DATE <- as.Date(pce$DATE)


timeseriesdata_inc_covid <- ts(pce$Month_12_Growth_Rate, start = c(2005, 01), end = c(2023, 04),
frequency = 12)
timeseriesdata_inc_covid

tsdisplay(timeseriesdata_inc_covid)

#------------------------------OVERALL PCE DATA ---------------------------
# Check for stationarity using ADF test
adf_test <- adf.test(timeseriesdata_inc_covid)
cat("ADF Test p-value:", adf_test$p.value, "\n")


# Split the time series data for the pre-covid period into training and testing sets using an 85-20 split
train_data_inc_covid <- subset(timeseriesdata_inc_covid , end = length(timeseriesdata_inc_covid)*0.85)
test_data_inc_covid <- subset(timeseriesdata_inc_covid , start =
((length(timeseriesdata_inc_covid)*0.85)+1))
head(train_data_inc_covid)
tail(train_data_inc_covid)
head(test_data_inc_covid)
tail(test_data_inc_covid)
summary(train_data_inc_covid)


#----------------------BASELINE ARIMA MODEL ------------------
```

```
arima_baseline <- auto.arima(train_data_inc_covid ,trace = TRUE,ic='aic')
#OVERALL PCE: ARIMA(1,0,3)(0,0,1)
#12 step ahead forecast
forecast_baseline_12 <- forecast(arima_baseline,h=12)
# Plot the forecasts and the actual values of the pre-covid test data
plot(forecast_baseline_12, main = "ARIMA Baseline Model 12 month-ahead Forecast")
lines(test_data_inc_covid, col = "red", lty = 1)
# Add a legend to the plot
legend("bottomright", legend = c("Forecast PCE", "Actual PCE"), col = c("black", "red"), lty = c(1, 1))

print(forecast_baseline_12)
# Calculate the accuracy metrics for the baseline model's forecasts
accuracy_baseline_12 <- accuracy(forecast_baseline_12, test_data_inc_covid[1:12,drop=FALSE])

# Print the accuracy metrics
print(accuracy_baseline_12)

checkresiduals(arima_baseline)
#CROSS VALIDATION
library(forecast)

# Specify the number of folds for cross-validation
num_folds <- 5

# Calculate the length of each fold
fold_length <- floor(length(timeseriesdata_inc_covid) / num_folds)

# Initialize a vector to store the cross-validation results
cv_results <- numeric(num_folds)

for (i in 1:num_folds) {
 # Define the start and end indices for the current fold
 start_idx <- (i - 1) * fold_length + 1
 end_idx <- start_idx + fold_length - 1

 # Split the data into training and validation sets
 fold_train <- timeseriesdata_inc_covid[-(start_idx:end_idx)]
 fold_valid <- timeseriesdata_inc_covid[start_idx:end_idx]

 # Fit the model on the training data
 model <- auto.arima(fold_train)

 # Forecast the next 'h' time steps with the correct length
 forecast <- forecast(model, h = length(fold_valid))

 # Calculate the accuracy of the forecast on the validation set
 accuracy <- accuracy(forecast, fold_valid)

 # Store the RMSE result in the cv_results vector
 cv_results[i] <- accuracy[1, "RMSE"]
```

```
}

# Print the cross-validation results
print(cv_results)

# Calculate the standard deviation of the original time series data
sd_original <- sd(timeseriesdata_inc_covid)

# Compare RMSE values to the standard deviation
print(cv_results / sd_original)

#IN-SAMPLE FORECASTING
residuals_baseline <- residuals(arima_baseline)
#Calculating the baseline R-squared value
baseline_r_squared <- 1 - (var(residuals_baseline) / var(train_data_inc_covid))
round(baseline_r_squared * 100, 3)

#OUT-OF-SAMPLE FORECASTING

test_data_inc_covid[1]
arima_baseline_out_1 <- Arima(test_data_inc_covid[1],model=arima_baseline)
accuracy(arima_baseline_out_1)

arima_baseline_out_3 <- Arima(test_data_inc_covid[1:3],model=arima_baseline)
accuracy(arima_baseline_out_3)

arima_baseline_out_6 <- Arima(test_data_inc_covid[1:6],model=arima_baseline)
accuracy(arima_baseline_out_6)

arima_baseline_out_9<- Arima(test_data_inc_covid[1:9],model=arima_baseline)
accuracy(arima_baseline_out_9)

arima_baseline_out_12 <- Arima(test_data_inc_covid[1:12],model=arima_baseline)
accuracy(arima_baseline_out_12)

arima_baseline_out_15 <- Arima(test_data_inc_covid[1:15],model=arima_baseline)
accuracy(arima_baseline_out_15)

#-------------------------------------------------------------------------
#----------------------MCSI -------------------------------------------
#--------------------------------------------------------------------------

# Load MCSI data from "mcsi_updated.csv" file (adjust the file path if needed)
mcsi_data <- read.csv("mcsi_updated.csv")

# Convert DATE column to Date format in MCSI data
mcsi_data$DATE <- as.Date(mcsi_data$DATE)


# Create time series data for MCSI in the pre-covid period
```

```
mcsi_timeseries_inc_covid <- ts(mcsi_data$Month_12_Growth_Rate, start = c(2005, 01), end = c(2023,
04), frequency = 12)
head(mcsi_timeseries_inc_covid)


# Split the time series data for MCSI in the pre-covid period into training and testing sets using an 85-20
split
train_mcsi_inc_covid <- subset(mcsi_timeseries_inc_covid, end = length(mcsi_timeseries_inc_covid) *
0.85)
head(train_mcsi_inc_covid)
test_mcsi_inc_covid <- subset(mcsi_timeseries_inc_covid, start = ((length(mcsi_timeseries_inc_covid) *
0.85) + 1))
head(test_mcsi_inc_covid)

# ----------------------ARIMAX Model for MCSI --------------------------------
length(train_data_inc_covid)
length(train_mcsi_inc_covid)
train_data_inc_covid
train_mcsi_inc_covid
# Fit the ARIMAX model using the pre-covid PCE data (train_data_inc_covid) and the MCSI data
(train_data_mcsi) as external regressors
arimax_mcsi <- auto.arima(train_data_inc_covid, xreg = train_mcsi_inc_covid, trace = TRUE, ic = 'aic')
checkresiduals(arimax_mcsi)
forecast_arimax_1 <- forecast(arimax_mcsi, xreg = test_mcsi_inc_covid[1:12])

plot(forecast_arimax_1, main = "ARIMAX 12 months ahead Forecast with MCSI Regressor")
lines(test_data_inc_covid, col = "red", lty = 1) # Plot actual PCE data in blue

# Add a legend to the plot
legend("bottomright", legend = c("Forecast MCSI", "Actual PCE"), col = c("black", "red"), lty = c(1, 1))

#Calculate and print the accuracy metrics for the forecast
accuracy_arimax_1 <- accuracy(forecast_arimax_1, test_data_inc_covid)
print(accuracy_arimax_1)

#IN-SAMPLE FORECASTING
# Get the residuals of the ARIMAX model
residuals_arimax_mcsi <- residuals(arimax_mcsi)

# Calculate the ARIMAX R-squared
arimax_r_squared_mcsi <- 1 - (var(residuals_arimax_mcsi) / var(train_data_inc_covid))
arimax_r_squared_mcsi

# Calculate incremental R-squared
incremental_r_squared_mcsi <- arimax_r_squared_mcsi - baseline_r_squared
round(incremental_r_squared_mcsi * 100, 3)

#OUT-OF-SAMPLE FORECASTING

test_mcsi_inc_covid[1]
```

```
arimax_mcsi_out_1 <- Arima(test_data_inc_covid[1],model=arimax_mcsi, matrix(test_mcsi_inc_covid[1],
nrow = 1, ncol = 1))
accuracy(arimax_mcsi_out_1)

arimax_mcsi_out_3 <-
Arima(test_data_inc_covid[1:3],model=arimax_mcsi,xreg=test_mcsi_inc_covid[1:3])
accuracy(arimax_mcsi_out_3)

arimax_mcsi_out_6 <-
Arima(test_data_inc_covid[1:6],model=arimax_mcsi,xreg=test_mcsi_inc_covid[1:6])
accuracy(arimax_mcsi_out_6)

arimax_mcsi_out_9<-
Arima(test_data_inc_covid[1:9],model=arimax_mcsi,xreg=test_mcsi_inc_covid[1:9])
accuracy(arimax_mcsi_out_9)

arimax_mcsi_out_12 <-
Arima(test_data_inc_covid[1:12],model=arimax_mcsi,xreg=test_mcsi_inc_covid[1:12])
accuracy(arimax_mcsi_out_12)

arimax_mcsi_out_15 <-
Arima(test_data_inc_covid[1:15],model=arimax_mcsi,xreg=test_mcsi_inc_covid[1:15])
accuracy(arimax_mcsi_out_15)

#--------------------------------------------------------------------
#----------------------CCI ------------------------------------
#-------------------------------------------------------------------
# Load CCI data from "cci_updated.csv" file (adjust the file path if needed)
cci_data <- read.csv("cci_updated.csv")

# Convert DATE column to Date format in CCI data
cci_data$DATE <- as.Date(cci_data$DATE)


# Create time series data for CCI in the pre-covid period
cci_timeseries_inc_covid <- ts(cci_data$Month_12_Growth_Rate, start = c(2005, 1), end = c(2023, 04),
frequency = 12)
head(cci_timeseries_inc_covid)

# Split the time series data for CCI in the pre-covid period into training and testing sets using an 85-20
split
train_cci_inc_covid <- subset(cci_timeseries_inc_covid, end = length(cci_timeseries_inc_covid) * 0.85)
head(train_cci_inc_covid)
test_cci_inc_covid <- subset(cci_timeseries_inc_covid, start = ((length(cci_timeseries_inc_covid) * 0.85) +
1))

# ---------------------ARIMAX Model for CCI -------------------------------
length(train_data_inc_covid)
length(train_cci_inc_covid)
```

```
# Fit the ARIMAX model using the pre-covid PCE data (train_data_inc_covid) and the CCI data
(train_cci_inc_covid) as external regressors
arimax_cci <- auto.arima(train_data_inc_covid, xreg = train_cci_inc_covid, trace = TRUE, ic = 'aic')
checkresiduals(arimax_cci)


forecast_arimax_1_cci <- forecast(arimax_cci, xreg = test_cci_inc_covid[1:12])

plot(forecast_arimax_1_cci, main = "ARIMAX 12 months Ahead Forecast with CCI Regressor")
lines(test_data_inc_covid, col = "red", lty = 1) # Plot actual PCE data in blue

# Add a legend to the plot
legend("bottomright", legend = c("Forecast", "Actual PCE"), col = c("black", "red"), lty = c(1, 1))

# Calculate and print the accuracy metrics for the forecast
accuracy_arimax_1_cci <- accuracy(forecast_arimax_1_cci, test_data_inc_covid)
print(accuracy_arimax_1_cci)

#IN-SAMPLE FORECASTNG
# Get the residuals of the ARIMAX model
residuals_arimax_cci <- residuals(arimax_cci)

# Calculate the ARIMAX R-squared
arimax_r_squared_cci <- 1 - (var(residuals_arimax_cci) / var(train_data_inc_covid))
arimax_r_squared_cci

# Calculate incremental R-squared
incremental_r_squared_cci <- arimax_r_squared_cci - baseline_r_squared
round(incremental_r_squared_cci * 100, 3)


#OUT-OF-SAMPLE FORECASTING

test_cci_inc_covid[1]
arimax_cci_out_1 <- Arima(test_data_inc_covid[1],model=arimax_cci, matrix(test_cci_inc_covid[1], nrow
= 1, ncol = 1))
accuracy(arimax_cci_out_1)

arimax_cci_out_3 <- Arima(test_data_inc_covid[1:3],model=arimax_cci,xreg=test_cci_inc_covid[1:3])
accuracy(arimax_cci_out_3)

arimax_cci_out_6 <- Arima(test_data_inc_covid[1:6],model=arimax_cci,xreg=test_cci_inc_covid[1:6])
accuracy(arimax_cci_out_6)

arimax_cci_out_9<- Arima(test_data_inc_covid[1:9],model=arimax_cci,xreg=test_cci_inc_covid[1:9])
accuracy(arimax_cci_out_9
arimax_cci_out_12 <- Arima(test_data_inc_covid[1:12],model=arimax_cci,xreg=test_cci_inc_covid[1:12])
accuracy(arimax_cci_out_12)

arimax_cci_out_15 <- Arima(test_data_inc_covid[1:15],model=arimax_cci,xreg=test_cci_inc_covid[1:15])
```

```
accuracy(arimax_cci_out_15)

#-------------------------------------------------------------------------------------
#--------------------Google Trends ------------------------------------------
#--------------------------------------------------------------------------------------

# Load data from file
google_data <- read.csv("pca_google_trends.csv")

# durable goods
 google_data <- read.csv("pca_durable.csv")

#non-durable goods
 google_data <- read.csv("pca_non_durable.csv")

#services
google_data <- read.csv("pca_service.csv")

# Create time series data for google in the pre-covid period
google_timeseries_inc_covid <- ts(google_data, start = c(2005,1), end = c(2023,04), frequency = 12)
head(google_timeseries_inc_covid)

training_len = round(0.85*nrow(google_timeseries_inc_covid))
training_len
testing_len = round(nrow(google_timeseries_inc_covid) - training_len)
testing_len
ts_split <- ts_split(ts.obj=google_timeseries_inc_covid, sample.out = testing_len)
train_google_inc_covid <- ts_split$train
length(train_google_inc_covid)
test_google_inc_covid <- ts_split$test
length(test_google_inc_covid)

# Fit the ARIMAX model
arimax_google <- auto.arima(train_data_inc_covid, xreg = train_google_inc_covid, trace = TRUE, ic = 'aic')
checkresiduals(arimax_google)

forecast_arimax_12_google <- forecast(arimax_google, xreg = test_google_inc_covid[1:12,,drop=FALSE])

plot(forecast_arimax_12_google, main = "ARIMAX 12 months Ahead Forecast with google trends
regressor")
lines(test_data_inc_covid, col = "red", lty = 1) # Plot actual PCE data in blue

# Add a legend to the plot
legend("bottomright", legend = c("Forecast", "Actual PCE"), col = c("black", "red"), lty = c(1, 1))

# Calculate and print the accuracy metrics for the forecast
accuracy_arimax_12_google <- accuracy(forecast_arimax_12_google, test_data_inc_covid)
print(accuracy_arimax_12_google)

#IN-SAMPLE FORECASTING
```

```
# Get the residuals of the ARIMAX model
residuals_arimax_gt <- residuals(arimax_google)

# Calculate the ARIMAX R-squared
arimax_r_squared_gt <- 1 - (var(residuals_arimax_gt) / var(train_data_inc_covid))
arimax_r_squared_gt

# Calculate incremental R-squared
incremental_r_squared_gt <- arimax_r_squared_gt - baseline_r_squared
round(incremental_r_squared_gt * 100, 3)

#OUT-OF-SAMPLE FORECASTING

test_google_inc_covid[1]
arimax_google_out_1 <- Arima(test_data_inc_covid[1],model=arimax_google,
xreg=test_google_inc_covid[1,,drop=FALSE])
accuracy(arimax_google_out_1)

arimax_google_out_3 <-
Arima(test_data_inc_covid[1:3],model=arimax_google,xreg=test_google_inc_covid[1:3,])
accuracy(arimax_google_out_3)

arimax_google_out_6 <-
Arima(test_data_inc_covid[1:6],model=arimax_google,xreg=test_google_inc_covid[1:6,])
accuracy(arimax_google_out_6)

arimax_google_out_9<-
Arima(test_data_inc_covid[1:9],model=arimax_google,xreg=test_google_inc_covid[1:9,])
accuracy(arimax_google_out_9)

arimax_google_out_12 <-
Arima(test_data_inc_covid[1:12],model=arimax_google,xreg=test_google_inc_covid[1:12,])
accuracy(arimax_google_out_12)

arimax_google_out_15 <-
Arima(test_data_inc_covid[1:15],model=arimax_google,xreg=test_google_inc_covid[1:15,])
accuracy(arimax_google_out_15)


#--------------------------------------------------------------------------------------
#--------------------Economic Policy Uncertainty ---------------------------
#---------------------------------------------------------------------------------------
# Load uncertainty data from "uncertainty_updated.csv" file
uncertainty_data <- read.csv("uncertainty_updated.csv")

# Convert DATE column to Date format in uncertainty data
uncertainty_data$DATE <- as.Date(uncertainty_data$DATE)

head(uncertainty_data)
# Create time series data for uncertainty in the pre-covid period
```

```
uncertainty_timeseries_inc_covid <- ts(uncertainty_data$Month_12_Growth_Rate, start = c(2005, 01),
end = c(2023, 04), frequency = 12)
uncertainty_timeseries_inc_covid

# Split the time series data for uncertainty in the pre-covid period into training and testing sets using an
85-20 split
train_uncertainty_inc_covid <- subset(uncertainty_timeseries_inc_covid, end =
length(uncertainty_timeseries_inc_covid) * 0.85)
head(train_uncertainty_inc_covid)
test_uncertainty_inc_covid <- subset(uncertainty_timeseries_inc_covid, start =
((length(uncertainty_timeseries_inc_covid) * 0.85)+1))
head(test_uncertainty_inc_covid)

# ----------------------ARIMAX Model for Uncertainty  --------------------------------
length(train_data_inc_covid)
length(train_uncertainty_inc_covid)
# Fit the ARIMAX model using the pre-covid PCE data (train_data_inc_covid) and the uncertainty data
(train_uncertainty_inc_covid) as external regressors
arimax_uncertainty <- auto.arima(train_data_inc_covid, xreg = train_uncertainty_inc_covid, trace = TRUE,
ic = 'aic')
checkresiduals(arimax_uncertainty)

forecast_arimax_1_uncertainty <- forecast(arimax_uncertainty, xreg = test_uncertainty_inc_covid[1:12])

plot(forecast_arimax_1_uncertainty, main = "ARIMAX 12 months Ahead Forecast with uncertainty
Regressor")
lines(test_data_inc_covid, col = "red", lty = 1) # Plot actual PCE data in blue

# Add a legend to the plot
legend("bottomright", legend = c("Forecast", "Actual PCE"), col = c("black", "red"), lty = c(1, 1))

# Calculate and print the accuracy metrics for the forecast
accuracy_arimax_1_uncertainty <- accuracy(forecast_arimax_1_uncertainty, test_data_inc_covid)
print(accuracy_arimax_1_uncertainty)

#IN-SAMPLE FORECASTING
# Get the residuals of the ARIMAX model
residuals_arimax_epu <- residuals(arimax_uncertainty)

# Calculate the ARIMAX R-squared
arimax_r_squared_epu <- 1 - (var(residuals_arimax_epu) / var(train_data_inc_covid))
arimax_r_squared_epu

# Calculate incremental R-squared
incremental_r_squared_epu <- arimax_r_squared_epu - baseline_r_squared
round(incremental_r_squared_epu * 100, 3)
incremental_r_squared_epu

#OUT-OF-SAMPLE FORECASTING
```

```
test_uncertainty_inc_covid[1]
arimax_uncertainty_out_1 <- Arima(test_data_inc_covid[1],model=arimax_uncertainty,
matrix(test_uncertainty_inc_covid[1], nrow = 1, ncol = 1))
accuracy(arimax_uncertainty_out_1)

arimax_uncertainty_out_3 <-
Arima(test_data_inc_covid[1:3],model=arimax_uncertainty,xreg=test_uncertainty_inc_covid[1:3])
accuracy(arimax_uncertainty_out_3)

arimax_uncertainty_out_6 <-
Arima(test_data_inc_covid[1:6],model=arimax_uncertainty,xreg=test_uncertainty_inc_covid[1:6])
accuracy(arimax_uncertainty_out_6)

arimax_uncertainty_out_9<-
Arima(test_data_inc_covid[1:9],model=arimax_uncertainty,xreg=test_uncertainty_inc_covid[1:9])
accuracy(arimax_uncertainty_out_9)

arimax_uncertainty_out_12 <-
Arima(test_data_inc_covid[1:12],model=arimax_uncertainty,xreg=test_uncertainty_inc_covid[1:12])
accuracy(arimax_uncertainty_out_12)

arimax_uncertainty_out_15 <-
Arima(test_data_inc_covid[1:15],model=arimax_uncertainty,xreg=test_uncertainty_inc_covid[1:15])
accuracy(arimax_uncertainty_out_15)


#-----------------------------------------------------------------------------------------------
#--------------------------------- Google + EPU -------------------------------------------
#-----------------------------------------------------------------------------------------------

train_google_epu_inc_covid <- cbind(train_uncertainty_inc_covid,train_google_inc_covid)
test_google_epu_inc_covid <- cbind(test_uncertainty_inc_covid, test_google_inc_covid)

head(train_google_epu_inc_covid)
head(test_google_epu_inc_covid)

# Make sure the column names are the same
colnames(test_google_epu_inc_covid) <- colnames(train_google_epu_inc_covid)

ncol(train_google_epu_inc_covid)
ncol(test_google_epu_inc_covid)
# Fit the ARIMAX model using the pre-covid PCE data (train_data_inc_covid) and the uncertainty data
(train_uncertainty_inc_covid) as external regressors
arimax_google_epu <- auto.arima(train_data_inc_covid, xreg = train_google_epu_inc_covid, trace =
TRUE, ic = 'aic')
checkresiduals(arimax_google_epu)


forecast_arimax_1_google_epu <- forecast(arimax_google_epu, xreg = test_google_epu_inc_covid[1:12,
])
```

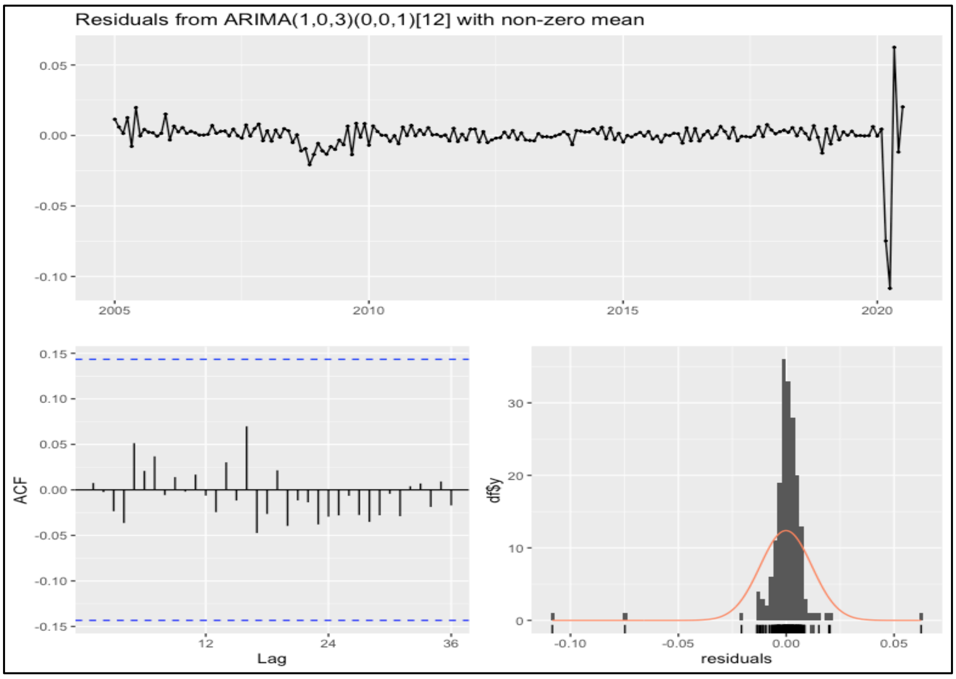
```
plot(forecast_arimax_1_google_epu, main = "ARIMAX 12 months Ahead Forecast with google and
uncertainty Regressor")
lines(test_data_inc_covid, col = "red", lty = 1) # Plot actual PCE data in red

#IN-SAMPLE FORECASTING
# Get the residuals of the ARIMAX model
residuals_arimax_gt_epu <- residuals(arimax_google_epu)

# Calculate the ARIMAX R-squared
arimax_r_squared_gt_epu <- 1 - (var(residuals_arimax_gt_epu) / var(train_data_inc_covid))
arimax_r_squared_gt_epu

# Calculate incremental R-squared
incremental_r_squared_gt_epu <- arimax_r_squared_gt_epu - baseline_r_squared
round(incremental_r_squared_gt_epu * 100, 3)
#OUT-OF-SAMPLE FORECASTING
arimax_google_epu_out_3 <-
Arima(test_data_inc_covid[1:3],model=arimax_google_epu,xreg=test_google_epu_inc_covid[1:3, ])
accuracy(arimax_google_epu_out_3)

arimax_google_epu_out_6 <-
Arima(test_data_inc_covid[1:6],model=arimax_google_epu,xreg=test_google_epu_inc_covid[1:6, ])
accuracy(arimax_google_epu_out_6)

arimax_google_epu_out_9 <-
Arima(test_data_inc_covid[1:9],model=arimax_google_epu,xreg=test_google_epu_inc_covid[1:9, ])
accuracy(arimax_google_epu_out_9)

arimax_google_epu_out_12 <-
Arima(test_data_inc_covid[1:12],model=arimax_google_epu,xreg=test_google_epu_inc_covid[1:12, ])
accuracy(arimax_google_epu_out_12)

arimax_google_epu_out_15 <-
Arima(test_data_inc_covid[1:15],model=arimax_google_epu,xreg=test_google_epu_inc_covid[1:15, ])
accuracy(arimax_google_epu_out_15)


#------------------------------------------------------------------------------------------------
#--------------------------------- Combination Model -------------------------------------------
#------------------------------------------------------------------------------------------------


train_combination_inc_covid <-
cbind(train_mcsi_inc_covid,train_cci_inc_covid,train_uncertainty_inc_covid,train_google_inc_covid)
test_combination_inc_covid <-
cbind(test_mcsi_inc_covid,test_cci_inc_covid,test_uncertainty_inc_covid,test_google_inc_covid)

head(train_combination_inc_covid)
head(test_combination_inc_covid)
```

```
# Make sure the column names are the same
colnames(test_combination_inc_covid) <- colnames(train_combination_inc_covid)

ncol(train_combination_inc_covid)
ncol(test_combination_inc_covid)
# Fit the ARIMAX model using the pre-covid PCE data (train_data_inc_covid) and the uncertainty data
(train_uncertainty_inc_covid) as external regressors
arimax_combination <- auto.arima(train_data_inc_covid, xreg = train_combination_inc_covid, trace =
TRUE, ic = 'aic')
checkresiduals(arimax_combination)


forecast_arimax_1_combination <- forecast(arimax_combination, xreg =
test_combination_inc_covid[1:12, ])

plot(forecast_arimax_1_combination, main = "ARIMAX 12 months Ahead Forecast with Combined
Model")
lines(test_data_inc_covid, col = "red", lty = 1) # Plot actual PCE data in red

#IN-SAMPLE FORECASTING
# Get the residuals of the ARIMAX model
residuals_arimax_com <- residuals(arimax_combination)

# Calculate the ARIMAX R-squared
arimax_r_squared_com <- 1 - (var(residuals_arimax_com) / var(train_data_inc_covid))
arimax_r_squared_com

# Calculate incremental R-squared
incremental_r_squared_com <- arimax_r_squared_com - baseline_r_squared
round(incremental_r_squared_com * 100, 3)


#OUT-OF-SAMPLE FORECASTING
arimax_combination_out_3 <-
Arima(test_data_inc_covid[1:3],model=arimax_combination,xreg=test_combination_inc_covid[1:3, ])
accuracy(arimax_combination_out_3)

arimax_combination_out_6 <-
Arima(test_data_inc_covid[1:6],model=arimax_combination,xreg=test_combination_inc_covid[1:6, ])
accuracy(arimax_combination_out_6)

arimax_combination_out_9 <-
Arima(test_data_inc_covid[1:9],model=arimax_combination,xreg=test_combination_inc_covid[1:9, ])
accuracy(arimax_combination_out_9)

arimax_combination_out_12 <-
Arima(test_data_inc_covid[1:12],model=arimax_combination,xreg=test_combination_inc_covid[1:12, ])
accuracy(arimax_combination_out_12)
```

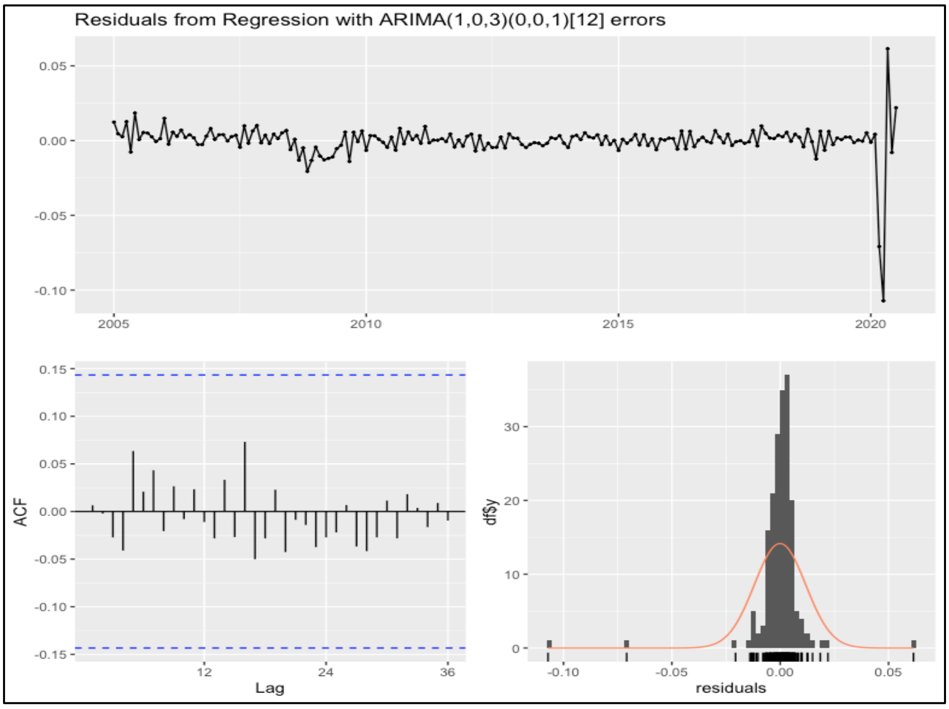
```
arimax_combination_out_15 <-
Arima(test_data_inc_covid[1:15],model=arimax_combination,xreg=test_combination_inc_covid[1:15, ])
accuracy(arimax_combination_out_15)


#---------------------------------------------------------------------------------------------
#----------------------------Diebold Mariano Test------------------------------
#---------------------------------------------------------------------------------------------
library(DMtest)

refit_baseline <- Arima(test_data_inc_covid,model=arima_baseline)
refit_mcsi <- Arima(test_data_inc_covid,model=arimax_mcsi,xreg=test_mcsi_inc_covid)
refit_cci <- Arima(test_data_inc_covid,model=arimax_cci,xreg=test_cci_inc_covid)
refit_google <- Arima(test_data_inc_covid,model=arimax_google,xreg=test_google_inc_covid)
refit_uncertainty <-
Arima(test_data_inc_covid,model=arimax_uncertainty,xreg=test_uncertainty_inc_covid)
refit_google_epu <-
Arima(test_data_inc_covid,model=arimax_google_epu,xreg=test_google_epu_inc_covid)
refit_combination <-
Arima(test_data_inc_covid,model=arimax_combination,xreg=test_combination_inc_covid)


# Create a list of refit models for each predictor variable
refit_models <- list(refit_mcsi, refit_cci, refit_google, refit_uncertainty, refit_google_epu,
refit_combination)
# List of predictor variable names
predictor_names <- c("MCSI", "CCI", "Google", "Uncertainty", "Google+EPU", "Combination")

horizons <- c(3, 6, 9, 12, 15)

results <- matrix(0, nrow = length(refit_models), ncol = length(horizons))
for (i in seq_along(refit_models)) {
  for (h in horizons) {
    dm_result <- dm.test(residuals(refit_baseline), residuals(refit_models[[i]]), h = h, alternative =
"greater")
    results[i, horizons == h] <- dm_result$p.value
  }
}
# Round the results to three decimal places
results <- round(results, 3)
# Create a data frame to store the results
results_df <- data.frame(Predictor = predictor_names, results)
# Rename the column names
colnames(results_df)[-1] <- paste("H =", horizons)
print(results_df)
write.csv(results_df, file = "dm_test_pces_inc.csv", row.names = FALSE)
```

7.2. **<u>Residual Plots of Models in the including COVID-19 period for Overall PCE</u>**

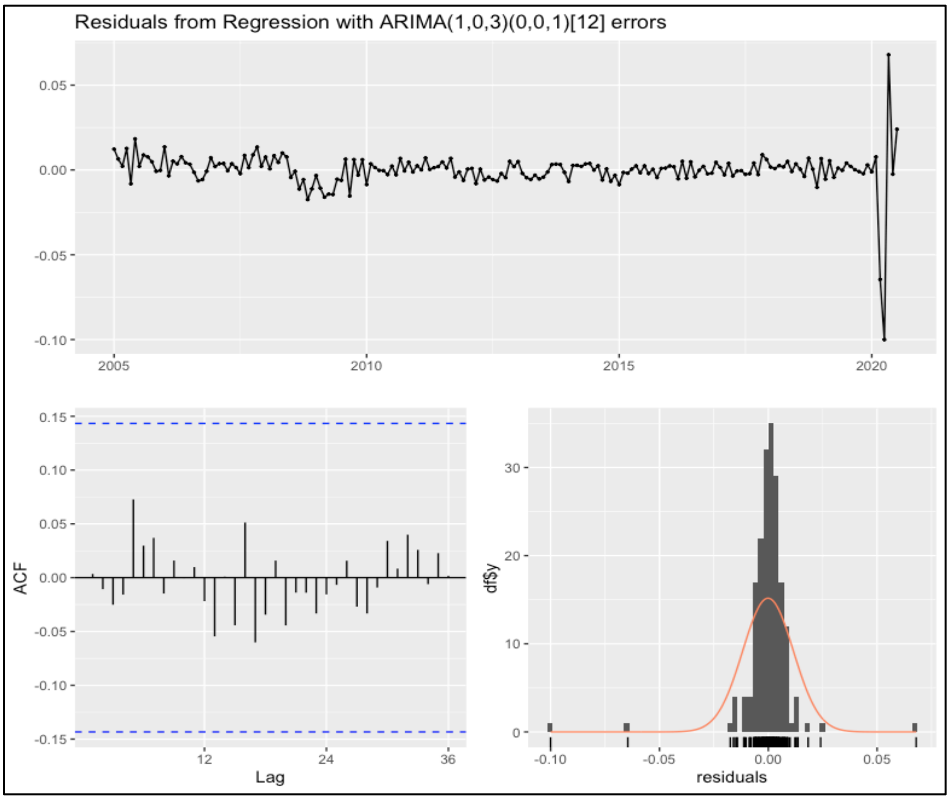
7.2.1. Baseline ARIMA model



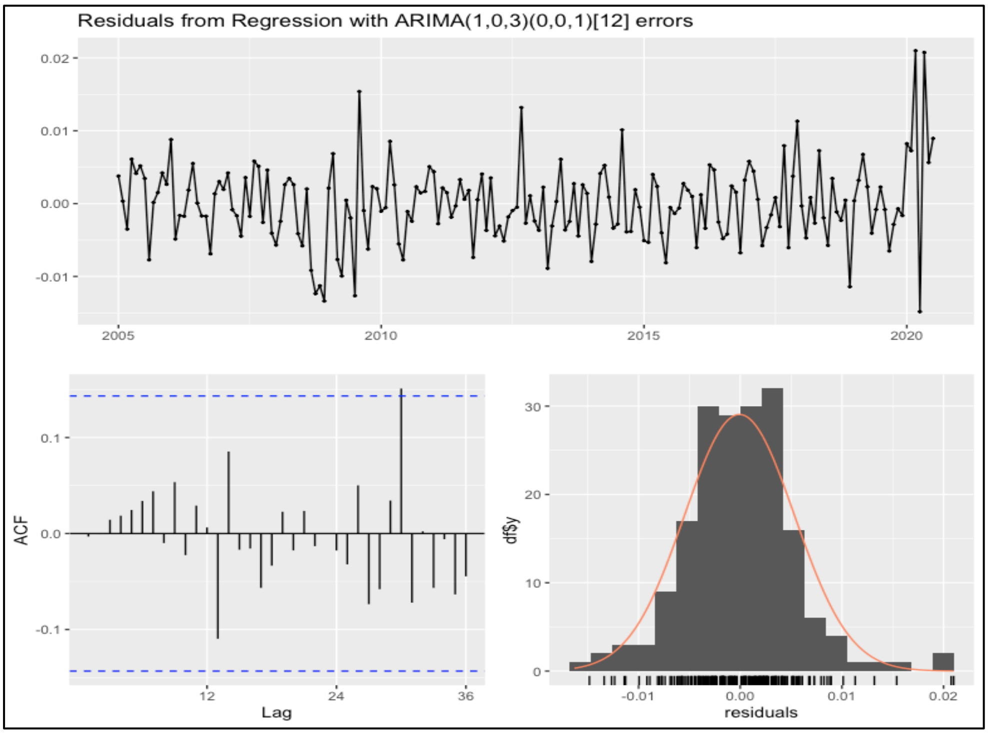
7.2.2. MCSI Model



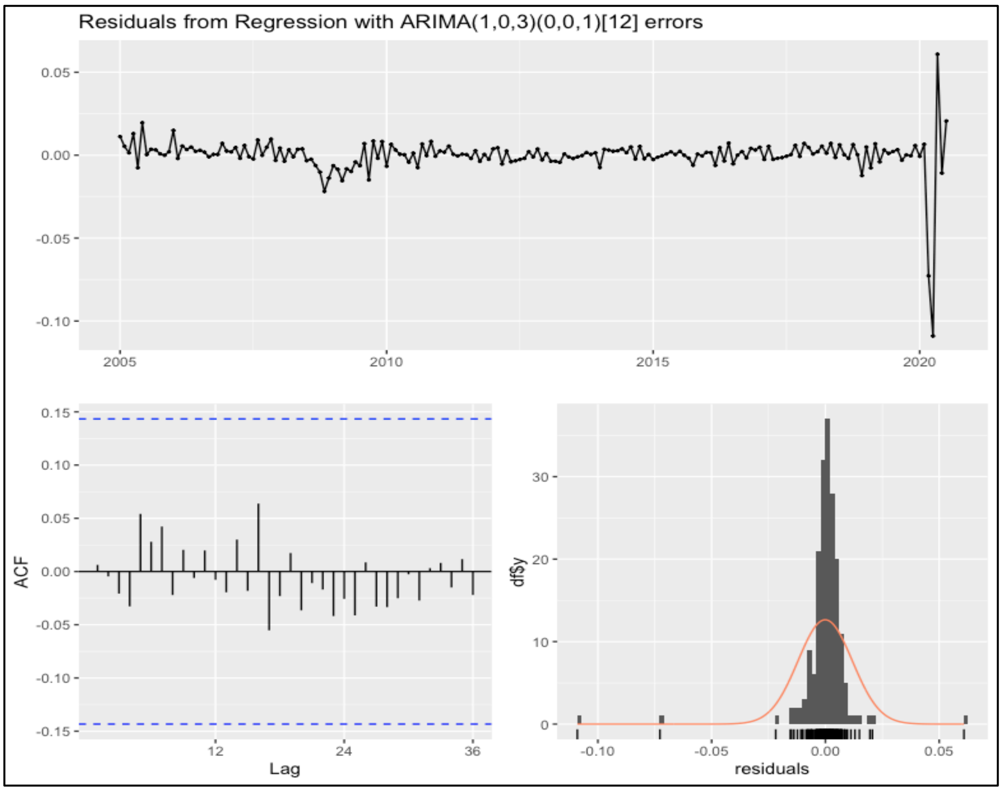
7.2.3. CCI Model


Residuals from Regression with ARIMA(1,0,3)(0,0,1)[12] errors
0.05
0.00
-0.05
-0.10
2005
2010
2015
2020
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
ACF
12
24
36
Lag
30
20
10
0
df$y
-0.10
-0.05
0.00
0.05
residuals

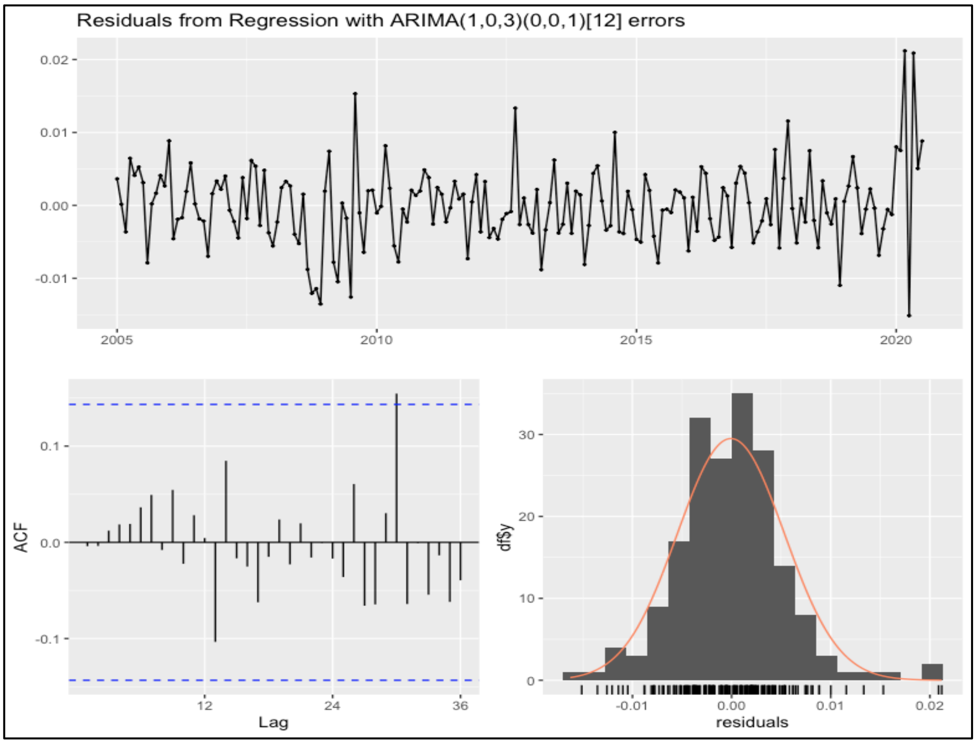
7.2.4. Google Trends Model



7.2.5. Economic Policy Uncertainty (EPU) Model



7.2.6. Google + EPU Model



7.2.7. Combination Model