



Faculty of Economics and Business

**Understanding the Intention to Consume by the Consumers for MSME
in Malaysia**

Lee Kim Seng

**DOCTOR OF BUSINESS ADMINISTRATION
2024**

Understanding the Intention to Consume by the Consumers for MSME
in Malaysia

Lee Kim Seng

A dissertation submitted
In fulfilment of the requirements for the degree of Doctor of Business Administration

Faculty of Economics and Business
UNIVERSITI MALAYSIA SARAWAK
2024

DECLARATION

I declare that the work in this dissertation was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgement has been given, it is the sole work of the author. The dissertation has not been accepted for, and is not being concurrently submitted as candidature for, any other degree.

.....

Signature

Name: Lee Kim Seng

Matric No.: 20040015

Faculty of Economics and Business

Universiti Malaysia Sarawak

Date : 29/4/2024

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who have directly or indirectly contributed to the creation of this dissertation.

I am sincerely grateful to the Centre for Graduate Studies for its invaluable guidance and unwavering support throughout my academic journey at Universiti Malaysia Sarawak.

Finally, my heartfelt gratitude goes to the university management for creating an environment that enabled me to successfully complete my studies here in Sarawak. To all those involved, I thank you for your contributions.

ABSTRACT

Micro, small and medium enterprises (MSMEs) make an important contribution to Malaysia's economy. The lack of market strategies is one of the biggest challenges for SMEs that hinders their growth. Private consumption as a measure of consumer spending that covers a multidisciplinary area is critical to economic development. The objective of this study is to investigate the impact of online search data on the estimation of private consumption in Malaysia. A comprehensive approach adopted incorporates both online and conventional models to estimate private consumption. The online search indicator used in the study was constructed from Google Trends data and focuses on product subcategories of the Shopee website in Malaysia. The dataset covers the period from 1 January 2016 to 31 March 2023. This research involves the development of an online search indicator for private consumption using principal component analysis, the construction of conventional leading economic indicators, and the estimation of online and conventional models for private consumption. The results of the variance decompositions suggest that the online model of private consumption could serve as a viable alternative or complement to the conventional model of private consumption. Looking ahead, this study envisages further exploration of search indicators for predicting private consumption, including multimodal models as data accumulation continues to increase.

Keywords: Google Trends, Private Consumption, Online Search, Malaysia

***Mentafsir Hasrat Pengguna dalam Pasaran MSME Malaysia:
Satu Analisis Komprehensif***

ABSTRAK

Perusahaan mikro, kecil dan sederhana memberi sumbangan penting kepada ekonomi Malaysia. Kekurangan strategi pasaran adalah salah satu cabaran terbesar bagi PKS dan menghalang pertumbuhan mereka. Penggunaan swasta sebagai ukuran perbelanjaan pengguna, kawasan pelbagai disiplin, adalah penting untuk pembangunan ekonomi. Objektif kajian ini adalah untuk menyiasat kesan data carian dalam talian terhadap anggaran penggunaan swasta di Malaysia. Pendekatan komprehensif diguna pakai yang menggabungkan kedua-dua model dalam talian dan konvensional untuk menganggarkan penggunaan swasta. Penunjuk carian dalam talian yang digunakan dalam kajian ini dibina daripada data Google Trends dan memfokuskan pada subkategori produk laman web Shopee di Malaysia. Set data meliputi tempoh dari 1 Januari 2016 hingga 31 Mac 2023. Penyelidikan ini melibatkan pembangunan penunjuk carian dalam talian untuk penggunaan swasta menggunakan analisis komponen utama, pembinaan penunjuk ekonomi utama konvensional dan anggaran model dalam talian dan konvensional untuk penggunaan swasta. Keputusan penguraian varians mencadangkan bahawa model penggunaan swasta dalam talian boleh berfungsi sebagai alternatif yang berdaya maju atau pelengkap kepada model penggunaan swasta konvensional. Memandang ke hadapan, kajian ini membayangkan penerokaan lanjut penunjuk carian untuk meramalkan penggunaan swasta, mungkin termasuk model multimodal apabila pengumpulan data terus meningkat.

Kata kunci: Google Trends, Penggunaan swasta, Pencarian dalam talian, Malaysia

TABLE OF CONTENTS

DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
ABSTRAK	iv
TABLE OF CONTENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	x
CHAPTER 1 INTRODUCTION	1
1.1 Study Background	1
1.2 Problem Statement	5
1.3 Research Questions	7
1.4 Objectives	8
1.5 Significance of the Study	10
1.6 Scope of the Study	11
1.7 Organisation of the Study	13
CHAPTER 2 LITERATURE REVIEW	15
2.1 Overview	15

2.2	Attention Economics Theory and AIDA Model	15
2.3	Online Search Data as Economic Activities	18
2.4	Private Consumption Using Google Trends Data	22
2.5	Consumer Private Consumption, MSMEs and Government Supports	27
2.6	Summary	29
CHAPTER 3 METHODOLOGY		31
3.1	Overview	31
3.2	Research Design	31
3.3	Data Collection	32
3.3.1	Online Search Data	32
3.3.2	Conventional Economic Data	34
3.3.3	Consumer Confidence Sentiments Index	36
3.3.4	Macroeconomic Variables	36
3.3.5	Sample Period	37
3.4	Data Analysis	38
3.4.1	Static Extraction of Online Search Indicators	39
3.4.2	Static Extraction of Conventional Economic Indicator	39
3.4.3	Online and Conventional Private Consumption Models	40
3.5	Research Framework	46
3.6	Summary	49

CHAPTER 4 DATA ANALYSIS AND FINDINGS	51
4.1 Overview	51
4.2 PCA Extraction of Online Search Indicator	51
4.2.1 Data Extraction from Google Trends	51
4.2.2 Results from Each Products Categories	53
4.2.3 Computation for Online Search Indicator	56
4.3 PCA Extraction of Conventional Economic Indicator	61
4.4 VAR Online and Conventional Model Analysis	64
4.4.1 Unit Root Test	64
4.4.2 Cointegration Test	68
4.4.3 VECM and Causality Test	73
4.4.4 VAR Variance Decomposition	80
4.5 Summary	87
CHAPTER 5 CONCLUSIONS	90
5.1 Summary of Findings	90
5.2 Implications of the Study	96
5.3 Limitations of the Study	97
5.4 Recommendations and Directions for Future Research	98
REFERENCES	101

LIST OF TABLES

Table 2.1: Number of Categories of COICOP 2018 by Level	23
Table 3.1: Data Construction for Conventional Economic Indicator	34
Table 3.2: Data Construction for Macroeconomic Variables	37
Table 3.3: Summary of Variables Used for Time Series Analysis	40
Table 4.1: Number of Keywords for Online Search Data	52
Table 4.2: Sample Adequacy for Online Search Data	54
Table 4.3: Reliability and Validity Measures for Online Search Data	55
Table 4.4: Total Variance Explained for Online Search Indicator	58
Table 4.5: Factor Loadings for Online Search Indicator	60
Table 4.6: Total Variance Explained for Conventional Economic Indicator	62
Table 4.7: Factor Loadings for Conventional Economic Indicator	63
Table 4.8: Unit Root and Stationarity Tests	66
Table 4.9: Cointegration Test – Online Private Consumption Model	69
Table 4.10: Cointegration Test – Conventional Private Consumption Model	71
Table 4.11: Granger Causality Results – Online Private Consumption Model	74
Table 4.12: Granger Causality Results – Conventional Private Consumption Model	77
Table 4.13: Variance Decomposition – Online Private Consumption Model	81
Table 4.14: Variance Decomposition – Conventional Private Consumption Model	84
Table 5.1: Summary of Methods and Results for Online Search Indicators and Conventional Economic Indicators	91

LIST OF FIGURES

Figure 2.1: AIDA Model	17
Figure 3.1: Research Framework	47
Figure 4.1: Short Run Causal Linkages – Online Private Consumption Model	76
Figure 4.2: Short Run Causality Direction – Conventional Private Consumption Model	79

LIST OF ABBREVIATIONS

ADF	Augmented Dickey-Fuller
AIDA	Attention, Interest, Desire and Action
AR	Autoregressive
BEA	Bureau of Economic Analysis
BNM	Bank Negara Malaysia
BVAR	Bayesian Vector Autoregressive
COICOP	Classification of Individual Consumption Purposes
DFGLS	Dickey-Fuller Generalised Least Squares
DOSM	Department of Statistics, Malaysia
ECM	Error Correction Model
ECT	Error Correction Term
GDP	Gross Domestic Product
ICT	Information and Communications Technology
KMO	Kaiser-Meyer-Olkin
MIER	Malaysian Institute of Economic Research
MSME	Micro, Small and Medium-sized Enterprise
NESDC	National Entrepreneur and SME Development Council
PCA	Principal Component Analysis

SVI	Search Volume Index
VAR	Vector Autoregressive
VECM	Vector Error Correction Model

CHAPTER 1

INTRODUCTION

1.1 Study Background

Malaysia is classified as an emerging market (International Monetary Fund, 2022, October; World Bank, 2021), and a country that is developing its economy and financial markets to a higher level. Malaysia has experienced considerable economic growth in recent decades (World Bank, 2022, November 29) and has made significant progress in developing its markets (Koen et al., 2017). However, Malaysia still faces challenges such as income inequality and political stability, which are typical characteristics of emerging economies.

Despite these challenges, small and medium-sized enterprises (SMEs) in Malaysia play a crucial role in the country's economy: They account for 97.4% of all enterprises and contribute 37.4% of Malaysia's GDP in 2021 (SME Corp., 2022), as reported by the National Entrepreneur and SME Development Council (NESDC) in MSME insights 2021. In addition, SMEs employ almost half (47.8%) of the country's workforce (SME Corp., 2022), highlighting their importance in economic growth and job creation.

However, MSMEs in Malaysia face a number of challenges and issues that can limit their growth and competitiveness. These include limited access to markets, finance, technology and innovation, human resource bottlenecks, limited management and technical skills, and regulatory barriers (Mohamad et al., 2021; Rahman et al., 2016; Tehseen, 2014). One of the biggest challenges for MSMEs in Malaysia is the lack of market knowledge

(Mohamad et al., 2021), which if sufficient, would help reduce uncertainty and support decision-making at all levels.

Market research enables businesses to better understand consumer behaviour, identify emerging trends, and develop effective marketing strategies; forecasting private consumption can also inform market research as it can help solve a range of economic and business problems, provide insights into consumer behaviour, reduce uncertainty and support decision-making at all levels. Private consumption forecasting is a multidisciplinary field that includes elements of economics, statistics, and data science.

Private consumption is an important element of GDP expenditure and acts as a key driver of the country's economic development that sustains the steady growth of the macroeconomy, which is of great importance for forecasting research. It is evident in 2022, when private consumption accounted for 56.7% (CEIC Data, 2023) of total expenditure in Malaysia.

Economists usually develop models to analyse and predict private consumption, drawing on theoretical foundations. They also use data on macroeconomic indicators such as GDP, inflation, and interest rates to make forecasts. Statisticians play a key role in developing and estimating models for forecasting private consumption, including time series models, structural models, and econometric models. They also use statistical techniques to assess the accuracy and reliability of forecasts. Data scientists apply machine learning algorithms to analyse large data sets, including big data sources such as social media, search engine queries, and transactional data, to improve the accuracy of private consumption forecasts.

Predicting private consumption in emerging economies is particularly challenging due to factors such as limited data, high volatility, and structural changes. There is motivation to explore the use of alternative data sources to overcome these challenges. The increasing amount of time Malaysians spend online and their preferences for certain Internet activities have made the Internet a valuable source for data mining. The Digital 2023 Global Overview Report (DataReportal, 2023a), published in collaboration with “We Are Social and Hootsuite”, reports that Malaysians spend an average of 8 hours and 6 minutes online per day, ranking 8th among countries with the highest online time. The 2021 Survey Report on ICT Use and Access by Individuals and Households, conducted by the Department of Statistics Malaysia (2023, May 31), showed that the top three activities in Internet use in 2021 were participating in social networking; downloading pictures, movies, videos or music; playing or downloading games; and searching for information on goods or services. In addition, the most common reason for using the Internet in Malaysia was to search for information, which 76.2% of users engaged in (DataReportal, 2023b). Google.com was the most visited website in Malaysia in February 2023, according to Semrush rankings, with nearly 1.3 billion monthly visits (2023), making Google Trends an accessible source of online search data. These findings suggest that online search data offers new opportunities for economic forecasting. By analysing the search terms used by consumers, it is possible to determine which products or services are in demand, and which can help businesses and policymakers to make informed decisions about production, marketing, and distribution. In addition, the use of online search data can help identify emerging trends in consumer behaviour, which can be used to anticipate changes in the market and adjust strategies accordingly.

As technology has advanced, people's lives have become increasingly intertwined with the Internet, leading to significant changes in the way people communicate and consume goods. This has led to the emergence of large amounts of online search data that can be used to predict private consumption. By analysing consumers' search behaviour in conjunction with their usage habits, attention generation will drive their search power and then potentially impact purchasing power. Insights into the factors that influence their purchasing decisions are gained through the AIDA (Attention, Interest, Desire, and Action) model; these are the basic stages that a consumer goes through when making a purchasing decision.

Decision models are frameworks that help explain how consumers make decisions, and they are often used in marketing research to develop more effective marketing strategies. The concept of decision models is also closely related to the theory of attention economics. Attention Economics Theory states that in a world where attention is a scarce resource, businesses and marketers must compete for consumer attention in order to be successful. Overall, Attention Economics Theory and decision models provide a framework for understanding how companies can compete for consumer attention in today's competitive markets. The integration of the AIDA model and the theory of attention economics can lead to the development of more comprehensive and accurate models for predicting private consumption, which can improve decision-making in various fields such as retail, advertising, and marketing.

1.2 Problem Statement

The challenges faced by MSMEs in achieving growth have long been an ongoing issue in Malaysia. In a study by Ambad et al (2020), based on interviews with business founders, marketing strategies are cited as one of the main obstacles to business growth in the country. Although MSMEs make up a significant 78.7% (SME Corp., 2022) of the total number of businesses, they often underestimate or overlook the strategic value of formal market research (Fuellhart & Glasmeier, 2003; Hutchinson et al., 2009; Lybaert, 1998; O'Donnell, 2011), relying instead on interpersonal relationships (Jocumsen, 2004). However, understanding customer demand is critical to gaining a competitive advantage and ensuring long-term sustainability. By utilising market research, MSMEs can improve their profitability, sustainability, and growth prospects. In addition, by acquiring comprehensive consumer data, MSMEs can tailor their products to the specific needs of customers, thereby aligning their marketing strategies with the company's growth objectives.

Standard official data are characterised by a slow flow of information and long publication times. Statistics on economic growth are usually published monthly or quarterly with a considerable delay. The standard forecasting approach involves waiting for all economic indicators to be collected, which can be time-consuming and limits the ability to make informed decisions in the short term. Furthermore, the use of lagged indicators can reduce the effectiveness of forecasts (Matsumoto et al., 2013) and influence the decisions of policy makers (Mustapa et al., 2019). Therefore, there is an urgent need for alternative data sources that can complement official published data for market research and provide a more up-to-date snapshot of the economic landscape.

Google Trends provides access to valuable Google search data that is used for forecasting purposes in various fields. These applications include predicting the outcome of political investigations (Mavragani & Tsagarakis, 2016), improving the forecasting of tourist flows and demand (Artola et al., 2015; Bangwayo-Skeete & Skeete, 2015; Choi & Varian, 2012; Jun et al., 2018), analysing the unemployment-related variables (Askitas & Zimmermann, 2009; Fondeur & Karamé, 2013; Vicente et al., 2015), assessing stock market investors' attention and trading decisions (Da et al., 2011; Moat et al., 2014; Preis et al., 2013), and examining the impact of the COVID-19 crisis (Costola et al., 2021; Shear et al., 2021) and private consumption (Kholodilin et al., 2010; Matsumoto et al., 2013; Vosen & Schmidt, 2011, 2012). Using these resources, MSMEs can better plan their marketing campaigns and focus on reaching their target audience efficiently.

These studies have been conducted in several advanced countries, including the United States (Kholodilin et al., 2010), the Netherlands (Veldhuizen et al., 2016), the Eurozone (Anttonen, 2018), Germany (Götz & Knetsch, 2019), and several countries such as Australia, Canada, the United Kingdom, and the United States (Perlin et al., 2017). However, there is limited research on the application of Google Trends data to understand and explore private consumption in the context of Malaysia, which is known as a multiracial country and an emerging market in Southeast Asia. Therefore, it is interesting to explore the use of online search data as an additional analytical indicator for forecasting private consumption in Malaysia.

1.3 Research Questions

Private consumption plays a crucial role in reflecting the buying habits and purchasing power of individuals (European Central Bank, 2018), making them an important customer base for SMEs. Consumer private consumption studies are crucial for retailers as they help them understand the needs, preferences, and attitudes of their customers (European Central Bank, 2018). By analysing consumer behaviour, retailers can identify trends and patterns that are useful for them to make informed decisions on product development, pricing, marketing, and other aspects of their business. This information can also help retailers tailor their offerings to the specific needs of different customer segments, which can lead to higher customer satisfaction and loyalty.

Overall, consumer research is an important tool for retailers who want to remain competitive in today's fast-paced retail environment. By understanding the needs and preferences of their customers, retailers can develop products and services that fulfil these needs, build strong relationships with their customers, and ultimately drive growth and profitability.

In the digital age, the emergence of new data sources has opened up new possibilities for estimating private consumption, making online search data particularly interesting for deciphering consumer behaviour in the field of marketing research. The aim of this study is therefore to investigate the impact of the inclusion of online search data on the estimation of private consumption. A comprehensive approach taken includes various empirical tests to assess the impact of both online and conventional models of private consumption. Particular attention is paid to explaining the estimation process of the private consumption model

involving online search data, with a focus on the Malaysian context. The specific research questions guiding this study are outlined below:

- i. What are the components of online search indicators? A major focus of this study is to identify and characterise the elements that make up online search indicators. In combination with conventional economic indicators, online search data can serve as a basic element for estimating private consumption.
- ii. What are the elements that make up conventional economic indicators? To ensure the effective use of private consumption estimates in the present and foreseeable future in Malaysia, the careful selection of leading economic indicators is of paramount importance. This question aims to identify and extract the components of conventional economic indicators that have the greatest impact on the accuracy of private consumption estimates.
- iii. Can the online model of private consumption serve as a viable alternative to conventional models for estimating private consumption? The objective is to explore the implications of the online private consumption model and the conventional private consumption model and to highlight their respective effectiveness and applicability in the Malaysian context.

1.4 Objectives

In addition to traditional methods of estimating private consumption, this study aims to use time series data of online search activities to complement the estimation of private

consumption. By incorporating the Consumer Confidence Sentiment Index to measure consumers' current market perceptions, the potential of online search data as a viable alternative for estimating private consumption is compared and evaluated. To achieve the main objective, the following specific objectives were set:

- i. To extract an online search indicator for private consumption using principal component analysis. This process involves the collection and analysis of consumer attention and search patterns on Google. The data will be extracted from Google Trends and processed to construct an online search indicator that effectively reflects consumer interest and behaviour in relation to private consumption.
- ii. To extract conventional economic indicators for private consumption through principal component analysis. Official data from reputable sources such as BNM, DOSM, CEICdata, and MIER are collected to identify key indicators for inclusion.
- iii. To estimate an online model of private consumption using online search indicators and a conventional model of private consumption based on established economic indicators. In both models, the Consumer Confidence Sentiments Index and macroeconomic variables such as gross domestic product and private consumption indicators were added. Various empirical tests, including unit root tests, cointegration tests, and vector error correction models, are used for a comprehensive analysis.

1.5 Significance of the Study

Private consumption forecasting plays a crucial role for various stakeholders such as businesses and governments as it provides valuable insights into the economy, consumer behaviour, and market trends.

For businesses, accurate forecasts of consumer spending can help with business planning as they can anticipate demand for their products or services. By analysing online search data, insights can be gained into the topics and products that attract consumer attention, which can influence marketing strategies and help businesses stay competitive. It can also help make informed decisions about investments in production, marketing and sales, and help investors predict market trends and make informed investment decisions.

For the government, private consumption is a crucial component of GDP, and therefore, accurate forecasts of consumer spending are important for economic policymakers. They can help identify various measures to boost the performance of MSMEs, including improving the institutional support system, facilitating access to finance, enhancing technological capabilities, improving market access, encouraging greater use of ICT, and raising awareness of product labelling and intellectual property rights protection.

The government can also take various measures to stimulate consumer spending. In 2021, for example, the government introduced the National Economic Recovery Plan (PENJANA) to stimulate the economy and support businesses and households affected by the COVID-19 pandemic. The plan includes measures such as cash gifts, tax incentives and discounts on the purchase of cars, houses, and other goods and services.

Overall, private consumption forecasting is crucial for businesses and policy makers as it helps in making informed decisions based on expected trends in consumer spending.

1.6 Scope of the Study

The main objective of this study is to investigate the potential of online search data as an alternative data source for predicting private consumption in addition to conventional economic indicators and to compare their effectiveness. Predicting private consumption is a critical aspect of economic forecasting as it has a significant impact on consumer behaviour and overall economic growth.

The data source is online search data that comes from the Google search engine, which records all searches conducted by online users. The Google search engine is the most popular search engine in Malaysia. Search data or keywords can be used to identify trends in consumer behaviour as it provides insight into what people are interested in and what they are looking for.

In the context of online search data, attention is important as users must first search for a specific keyword before they can access the information they are interested in. Therefore, this study used the AIDA model and the theory of attention economics to gain insights into the components that influence consumers' purchase decisions. The AIDA model provides a deeper understanding of consumer behaviour in predicting private consumption, which can support the development of more effective marketing strategies; while the Attention Economy Theory can provide insight into how online search data can be used to

identify patterns of consumer behaviour and detect changes in consumer preferences and needs.

The online search data used in the study is sourced from Google Trends, which includes all searches conducted by online users via Google Search. The keywords used to download the search data are taken from the category names of Shopee websites in Malaysia. However, it should be noted that the study may not take into account all the factors that influence private consumption. Other economic, social, and political components that could affect private consumption may not be included in the online search data.

The study is limited to the availability and quality of search data and is conducted within a specific time period. Data for all searches conducted in Malaysia using Google Search can be downloaded from Google Trends from January 2004, when Google Trends data first became available. The study will use monthly time series data in Malaysia for the period from January 2016 to March 2023. It is important to note that the study may not be generalisable across different time periods and geographical regions.

The study will use quantitative research methods, statistical econometric analysis (PCA), and time series forecasting analysis (VAR) to test the predicted effectiveness. Qualitative research methods will not be considered in this study. Overall, the results of this study will provide up-to-date marketing information that will help reduce uncertainty and enable informed decisions for MSMEs.

1.7 Organisation of the Study

This study is divided into five chapters, which includes introduction, literature review, research methodology, data analysis, and conclusion. The following is a brief description of the information related to each chapter.

Introduction in Chapter 1 outlines the scope of this study, the reasons for conducting this study, the importance of this topic, in addition to explaining the current scenario and the research questions regarding the objectives, rationale, and contributions of this study.

The literature review in Chapter 2 provides a comprehensive overview of the theory and literature on the use of online search data for forecasting private consumption. A thorough review of previous studies conducted in different geographical regions is conducted to gain insight into the existing academic and commercial publications on private consumption forecasting, conventional economic indicators and the use of online search data in economic analysis. The first section of this chapter reviews previous research on the use of online search data, including Google Trends, for forecasting. The following sections focus on forecasting private consumption, drawing on both theoretical and empirical evidence to support the conceptual framework and assumptions of this study.

Chapter 3 emphasises on a methodology for collecting data of this study, coming from various online sources, including Google Trends, to gain insight into online search behaviour in the context of private consumption. It provides an overview of the structure of Google Trends data and outlines the steps to obtain appropriate time series data for analysis. The chapter discusses techniques for creating indicators for online search data and traditional economic indicators. These indicators are used to build an integrated model of private

consumption that is tested for effectiveness. In addition, the statistical methods used in this study, including PCA and time series analysis, are explained.

The data analysis in Chapter 4 is used to develop online search indicators from online search data and to develop conventional economic indicators from conventional economic data. Each of these indicators is integrated with the consumer confidence sentiment index and macroeconomic variables to build an online private consumption model and a conventional private consumption model. The data analysis methods used in this study include unit root test, VAR or cointegration test, causality test, VECM, and variance decomposition to compare the effectiveness of both models.

Chapter 5 is the conclusion; the results of the study are summarised and the effectiveness of using online search data as an alternative analytical indicator for forecasting private consumption is discussed. The conclusions of this study and its contribution to practise are highlighted. The limitations and implications of the study are also discussed, along with suggestions for further research directions.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

This chapter mainly reviews the previous literature on the theoretical and empirical aspects of private consumption forecasting, especially based on online search data. After this overview, the definition of the theory, Google search, and online search data will be started to find out the popularity of this topic. Then, Google Trends or online search articles dealing with forecasting will be selected for in-depth study. Articles on forecasting private consumption are then discussed and how online search data can be an alternative data source to complement the conventional economic data.

2.2 Attention Economics Theory and AIDA Model

The theory behind the idea of “producers compete for consumer attention rather than consumer spending” is based on the concept of "attention economics"

Attention economics is a theory first introduced in the 1970s by the psychologist Simon (1971). The theory is based on the idea that attention is a scarce resource and that individuals and organisations must compete for it to be successful. In other words, attention is considered a valuable commodity, just like money or time, and companies and marketers must find ways to capture and hold consumers' attention in order to be successful.

According to attention economics, a potential customer's attention is treated as a resource (Pedrycz & Chen, 2013); individuals have a limited amount of attention to allocate to various stimuli such as products, services, and marketing messages. This means that businesses and marketers need to find ways to stand out from the competition and capture consumers' attention to succeed. This can be done in a variety of ways, such as creating engaging content, developing eye-catching visuals, or offering unique products or services.

The concept of decision models is closely related to attention economics. Decision models are frameworks that help explain how individuals make decisions, and they are often used in marketing research to develop more effective marketing strategies. There are different types of decision models, but they all share the common goal of helping companies and marketers understand how consumers make decisions and what factors influence those decisions.

The "Consumer Decision-Making Process" is a commonly used decision model. Lee and Hoffman (2015) divide the purchase decision process into several phases, namely attention, interest, desire, and action. The AIDA model summarises the main phases a consumer goes through when making a purchase decision:

- Attention: Grab the consumer's attention through marketing messages, advertisements or other promotional materials.
- Interest: After gaining the consumer's attention, arouse their interest in your product or service by highlighting its features, benefits and unique selling points.

- **Desire:** Once interest is aroused, create desire for your product or service by showing how it can solve the consumer's problem or satisfy their needs.
- **Act:** Finally, prompt the consumer to take action and make the purchase by giving a clear call to action, e.g. "Buy now" or "Sign up today"

Due to its straightforward approach, the AIDA model is more popular than the five-step model. Bartels and Onwezen (2014) divide the purchase decision into five stages, including problem recognition, information search, evaluation of alternatives, purchase strategy, and post-purchase behaviour. This model, known as Kotler's five-stage model, explicitly shows how a company can improve its product attributes in response to growing consumer demand.

As illustrated in Figure 2.1, while the AIDA model is a simplified version of the consumer decision-making process, it still provides a useful framework for companies and marketers to understand how to effectively communicate with consumers and influence their behaviour. By understanding how consumers go through these stages and what factors influence their decisions, businesses can develop more effective marketing strategies and better compete for consumers' attention.

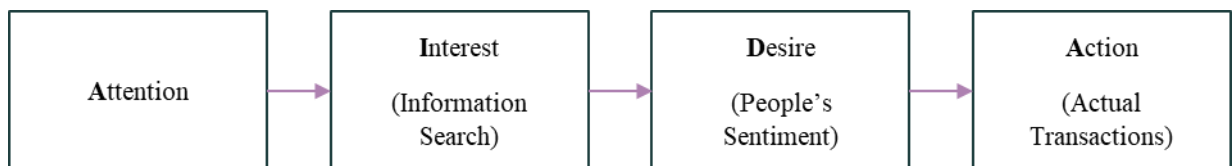


Figure 2.1: AIDA Model

Looking at the concepts of attention economics and decision models, it is clear that consumers of the Internet generation are driven by attention, which can affect their purchasing power. This is due to the fact that attention is a valuable resource in the modern world where individuals must compete to achieve their goal.

In addition, the way people communicate and interact with each other has changed significantly in the age of information networks. Instead of relying on face-to-face communication or traditional means of communication, people can now brainstorm and discuss on the Internet. This shift in communication methods has led to a huge amount of information being transmitted in natural language over the Internet, increasing the importance of attention in the modern world.

2.3 Online Search Data as Economic Activities

Da, Engelberg, and Gao (2011) investigated the concept of search attention and introduced the Search Volume Index (SVI) as a measure of the degree of attention people pay to search engines. By analysing search volume for stocks between 2004 and 2008, they compared the SVI with existing indicators of investment focus, such as the IPO Index and the VIX Fear Index, and identified three key findings. First, the SVI showed different correlations with the existing indicators. Second, the SVI was found to be more sensitive to changes in investor attention. Finally, it was particularly effective in measuring retail investor attention. The study found that an increase in the SVI led to an increase in stock prices over the next two weeks, but with the possibility of a reversal over the next year. Moreover, an increase in SVI also led to higher IPO returns on the first day, but relatively

lower long-term stock price returns. Hamid and Heiden (2015) also used search attention to develop weekly indicators to predict investor behaviour in the stock market.

Google Trends provides time series data on the frequency of popularity of search queries via Google Search. The data has been made available for free of charge since 2004. Users can specify geographic locations for a specific time period and enter search terms in any language to perform a search in the Google Trends search box. Google Trends can exclude irrelevant search queries by specifying the desired keyword. Google Trends outputs the Search Volume Index (SVI), the trend in popularity of a particular search relative to time and place over a period of time, where 100 means the highest popularity. SVI values are the relative popularity in relation to the total number of searches per location and time period. Normalisation depends on the specific search terms, so SVI values from different locations cannot be compared.

Currently, many studies use data from various public domains, such as search engines and social networks, to conduct research on various topics. Choi and Varian (2012) used data from the Google search engine to predict the value of economic indicators such as car sales, unemployment benefits, travel destination planning, and consumer confidence. Although the government regularly publishes economic indicators, statistical information is often outdated. Consequently, there is a need for predictive capabilities for these economic indicators. Tools such as Google Trends, UPS, Mastercard, and others can provide real-time data from which valuable information can be gleaned, and this is in line with Da, Engelberg, and Gao (2011) when they state, "search will bring attention".

For example, to assess the impact of car sales, Choi and Varian (2012) used data from Google search keywords such as recreational vehicles, car insurance and trucks. Using a seasonal AR model with an estimated period from January 2004 to July 2011, they found that the number of search terms significantly affects car sales. When this method is applied to topics such as unemployment benefits, tourism destinations, and consumer confidence, the explanatory power improves compared to the traditional model, with all coefficients of determination increasing. Overall, the inclusion of Google keywords in the model improved performance by 5 to 20% compared to the simple seasonal model AR model.

To predict the real estate and automotive market in Hungary, Toth and Hajdu (2012) used Google Trends search series to identify relevant indicators. They also investigated whether the Google Trends search sequence is only applicable in large European and American countries with advanced Internet technology or whether it can also be used in countries with relatively low Internet penetration. The results showed that the Trends index works well in predicting the real estate and automotive markets in Hungary, suggesting that it is also applicable in countries with low Internet penetration. Moreover, the short-term predictive power of the Trends Index is better than that of economic indicators that rely only on backward-looking information from the past.

Herzog (2014) conducted a study using Google search data to analyse the impact on corporate asset prices. The study covered 30 international banks from 2004 to 2010 and a time series model was constructed. The study produced two main findings. First, when evaluating the impact of Google search volume and Google search clicks on corporate asset prices, it was found that the rate of change of Google search volume had a positive relationship with the price of a company's assets. On the other hand, Google search clicks

were positively related to the trading volume of a firm's shares. Second, in terms of the dynamic relationship of assets, Google search volume and Google search clicks were found to have different effects.

Pavlicek (2014) looked at Google econometrics and examined the impact of job-related Google search terms on the unemployment rate. The study used two models called autoregressive (AR) and VAR, to analyse the data. The analysis focussed on the Czech Republic and found that Google search terms served as explanatory factors. The model VAR was used to perform a causal analysis, a shock response, and a disassembly of the prediction error. The results supported the hypothesis that Google Trends have an impact on the unemployment rate. Furthermore, Pavlicek (2014) suggested that the inclusion of Google search data in the analytical model improves the accuracy of the predictive power compared to traditional variable models.

According to Boik, Greenstein, and Prince (2017), manufacturers in today's competitive markets no longer compete directly for consumer spending, but focus on consumer attention. In their study, they examined how households use online tools to focus their attention on the markets they are interested in and how they then allocate their attention. Households' attention allocation decisions can be divided into three main directions, namely allocation quantity, allocation location, and allocation type. The study looked at click traffic data from 2008 to 2013, which shows that the amount of time households spend searching online has increased significantly. Using chats, videos, social media and other factors, the study found that the large number of expected changes indicated the consumption status of the household sector. The study also found that with higher household income, the flexibility

of attention is greater and the stability of dispersion at different network intercept points is similar, while the density of attention on websites is quite stable.

In addition, the study pointed out that although some studies suggest that income and other economic variables can affect the total amount of time people spend on the Internet, the magnitude and depth of changes that can be predicted in demographic data tend to be smaller when demographic patterns transition. People's income is not affected by family education, family size, age, and whether there are children using and paying attention to the Internet. The study highlights that with the development of various types of network device software, compared to 2008, the public has paid more attention to online services and social media software.

2.4 Private Consumption Using Google Trends Data

Kholodilin, Podstawski, and Siliverstovs (2010) evaluated the effectiveness of nowcasting private consumption in the US using Google Trends search indicators. Their evaluation model incorporated general economic variables, financial indicators and Internet search indicators. To determine the search indicators, they used bulk searches for consumer goods and extracted the top 10 search terms on Google. The results of the study suggest that the private consumption forecasting model included in the search index has higher predictive power than previous short-term forecasting models.

Vosen and Schmidt (2011) conducted a time series analysis of private consumption using Google Trends. They found that using Google Trends as an indicator provides better

predictions of in-sample and out-of-sample consumer behaviour than traditional questionnaires such as the Consumer Sentiment Index and the Consumer Confidence Index. The authors proposed to include Google Trends data as an explanatory variable in the prediction of private consumption. To facilitate this, they organised the relevant items in the Bureau of Economic Analysis (BEA) and Google classification. As this study is conducted in Malaysia, the goods and services are based on the Classification of Individual Consumption Purposes (COICOP), year 2018. Table 2.1 lists all the first level categories of consumption expenditure defined in COICOP (United Nations, 2018), it consists of 15 divisions, 63 groups and 183 classes. In line with the public term, the consideration of the online search data to be sourced for this study also corresponds to the COICOP 2018 classification.

Table 2.1: Number of Categories of COICOP 2018 by Level

Code	Description	#	#	#
		Division	Group	Class
01	Food and non-alcoholic beverages	1	3	17
02	Alcoholic beverages, tobacco and narcotics	1	4	7
03	Clothing and footwear	1	2	6
04	Housing, water, electricity, gas and other fuels	1	5	15
05	Furnishings, household equipment and routine household maintenance	1	6	13
06	Health	1	4	11
07	Transport	1	4	16
08	Information and communication	1	3	13
09	Recreation, sport and culture	1	8	24
10	Education services	1	5	5
11	Restaurants and accommodation services	1	2	3
12	Insurance and financial services	1	2	8

13	Personal care, social protection and miscellaneous goods and services	1	4	8
14	Individual consumption expenditure of non-profit institutions serving households (NPISHS)	1	6	22
15	Individual consumption expenditure of general government	1	5	18
Total		15	63	186

Note. Author created for this study

Division, is the total no. of Division (or two-digit level)

Group, is the total no. of Group (or three-digit level)

Class, is the total no. of Class (or four-digit level)

Source: Compiled for this study, adopted from COICOP (United Nations, 2018)

With private consumption accounting for about 70% of total spending in the United States, real-time information on consumer behaviour is crucial. However, official data are usually published with a one-month delay. Leading indicators and high-frequency data can help predict future trends and provide insights into the current situation, making the release of high-frequency data particularly useful in times of great uncertainty and major events.

Traditionally, sentiment indicators are based on questionnaires that take into account economic and psychological aspects. However, the empirical literature has shown that consumer sentiment indicators in the United States are closely related to consumer behaviour. In predicting private consumption, the questionnaire-based indicator is consumers' propensity to buy, while Google Trends provides a preliminary measure of consumers' preparation for their next purchase.

Vosen and Schmidt (2011) used a simple autoregressive model to forecast private consumption growth. They found that adding Google Trends data to the model improved in-

sample and out-of-sample predictions. To improve the accuracy of the analysis, the authors recommended using at least two years of data and seasonally adjusting the data. Overall, the study suggests that Google Trends is a valuable source of real-time information for predicting private consumption.

Vosen and Schmidt (2012) used Google's Big Data to predict private consumption in Germany and found that this new type of indicator was more accurate than traditional predictive indicators. They created monthly consumption indicators and used autoregressive models to make predictions.

Matsumoto, Matsumura, and Shiraki (2013) demonstrated the problem of the timeliness of official Japanese data using the example of the earthquake in Japan. Official data on new vehicle registrations, economic surveillance surveys, commercial and industrial production surveys, service sector surveys, tourism enterprise revenue surveys, and quarterly GDP estimates are all lagged by different periods, making them outdated for timely analysis. To test the representativeness of the search attention index, they conducted a survey on tourism using Google Trends search sequences on 11 tourism categories, including characteristic tourism, destinations, travel records, and tourism offices. After creating the series and dividing the data into three factors by PCA, a forecast was made after seasonal adjustment.

The empirical results showed that the first principal component factor had an overall explanatory power of 0.08 on tourism expenditure, and the individual coefficients had a very significant positive impact with a coefficient value of 0.61. After adding survey data, the

overall predictive power was higher than when using economic forecast questionnaires alone, demonstrating the representativeness of the Google Trends metrics.

The current focus of the literature is to explore how consumer confidence can improve the accuracy of household consumption forecasts in different countries around the world. Gausden and Hasan (2018) examined how the inclusion of consumer confidence in the UK can lead to better forecasts of household consumption. Similarly, Juhro, and Lyke (2020) examined how consumer confidence can help forecast household consumption in Indonesia and found that policymakers can improve their forecast accuracy by 4 to 13 percentage points by including consumer sentiment and business sentiment in their forecasting models. This strand of literature consistently suggests that the use of survey-based indicators such as consumer confidence or sentiment improves the forecasting accuracy of private consumption. In this empirical analysis, the consumer confidence index is included as a predictor to compare the predictive power of Internet search volume data with consumer confidence.

Woo and Owen (2019) considered survey-based indicators as a complement rather than a replacement for Google Trends data. They evaluated forecasts for durable goods, non-durable goods and services separately, and also analysed the impact of including Google Trends data related to news and consumption separately. Their results suggest that Google Trends data increases the accuracy of forecasts for all consumption sectors, although the degree of improvement varies by model specification.

2.5 Consumer Private Consumption, MSMEs and Government Supports

In terms of business size, most SMEs were micro enterprises, accounting for 78.7% of the total SMEs in Malaysia (SME Corp., 2022). Private consumption studies are crucial for MSMEs as they help them understand the needs, preferences and attitudes of their customers. By analysing consumer behaviour, MSMEs can identify trends and patterns that can help them make informed decisions on product development, pricing, marketing and other aspects of their business. This information can also help MSMEs tailor their offerings to the specific needs of different customer segments, which can lead to higher customer satisfaction and loyalty.

Overall, consumer research is an important tool for MSMEs that want to remain competitive in today's fast-paced retail environment. By understanding the needs and preferences of their customers, MSMEs can develop products and services that fulfil these needs, build strong relationships with their customers and ultimately drive growth and profitability.

The Malaysian government has taken a number of measures to support MSMEs and private consumption as it recognises their key role in promoting economic growth and development. Below are the key initiatives to support MSMEs.

SME Bank Malaysia provides financial support to SMEs through various lending and financing programmes that focus on areas such as technology adoption and green initiatives. Government grants such as the Microcredit Scheme and the High-Impact Project Grant provide financial support to micro and small enterprises for specific projects or upgrading of facilities. For capacity building, SME Corporation Malaysia (SMECorp) is the

national agency that provides various training programmes, workshops, and mentorship initiatives to improve entrepreneurial skills, digital usage, and market access for SMEs. Malaysia Digital Economy Corporation (MDEC) focusses on promoting digital adoption and technology entrepreneurship among SMEs through programmes such as the National SME Digitalisation Grant and talent development initiatives.

In addition, efforts are currently being made to simplify procedures for business registration, authorisations, and tax declarations in order to reduce the administrative burden on MSMEs. Flexible or customised schemes for MSMEs are used to address their specific needs and constraints. Sales Tax Exemptions on essential goods and services such as basic foodstuffs and public transport help to reduce the cost of living and encourage spending. Digital initiatives, particularly e-commerce programmes and campaigns, encourage online shopping and cashless transactions, facilitating convenient and accessible consumption options. Cybersecurity measures, by building trust and ensuring online safety through robust cybersecurity measures, promote consumer confidence in e-commerce and digital transactions.

It is important to note that these are just some of the key measures and that specific details and effectiveness may vary as they depend on the programme and economic climate. The Malaysian government regularly reviews and adapts its policies to address new challenges and support both MSMEs and private consumption for sustainable economic growth.

2.6 Summary

A comprehensive review of the literature revolves around the prediction of private consumption, with a particular focus on the use of online search data, especially Google Trends. The chapter begins with an overview of the theoretical and empirical aspects of private consumption forecasting, emphasising the growing importance of online search data as an alternative data source.

The theory of attention economics and the AIDA model were discussed to provide a theoretical basis for understanding that attention is a valuable resource in the modern world. The AIDA model, which represents the consumer's decision-making process, emphasises the stages of attention, interest, desire, and action. This theoretical framework formed the basis for understanding the role of attention in shaping consumer behaviour.

Section 2.3 on online search data as an economic activity examined studies that analysed the relationship between search attention, Google Trends data, and economic indicators. Various research findings showed the potential of using online search data for nowcasting and predicting economic activity, including private consumption.

The subsequent Section 2.4 focussed on private consumption using Google Trends data, with studies demonstrating the effectiveness of incorporating Google Trends into estimation models. The research showed that Google Trends data can improve the predictive power of personal consumption growth models and provide valuable online insights.

In addition, this chapter has emphasised the importance of considering consumer confidence alongside online search data in order to obtain more accurate forecasts of private

consumption. Studies have shown that the inclusion of survey-based indicators improves forecasting accuracy, with Google Trends data complementing traditional approaches.

Finally, the importance of private consumption for micro, small, and medium enterprises and the Malaysian government's initiatives to support MSMEs and private consumption were discussed. Various programmes, grants, and capacity building initiatives were presented to illustrate the government's commitment in promoting the growth of MSMEs and private consumption.

This chapter laid the foundation for the empirical analysis in the following chapters. The theoretical framework, the studies on attention economics, and the empirical evidence on the use of Google Trends data to estimate private consumption provide a solid foundation for understanding the dynamics of consumer behaviour in the digital age and the potential of online data for economic forecasting.

CHAPTER 3

METHODOLOGY

3.1 Overview

This chapter provides an overview of the research paradigm, philosophy, methodology, and methods used in the study. It outlines the data collection process, including the collection of online search data and the sourcing of published data for conventional economic indicators, the consumer confidence sentiment index, and macroeconomic variables. It then explains the statistical techniques used in data analysis such as PCA and VAR. Finally, the statistical methods used in the study are explained in detail.

3.2 Research Design

The research paradigm of this study is quantitative research with an empirical and exploratory approach. It adopts a positivist epistemological approach, which involves the collection and analysis of numerical data from Google Trends and traditional survey-based economic data to predict private consumption. The study uses time series analysis, a common research method for studying the behaviour of variables over time. In this study, time series analysis is used to analyse the relationship between Google Trends data and traditional economic indicators and their ability to predict private consumption. Time series analysis analyses historical data in order to make predictions about future trends or

behaviours. The study uses VAR to analyse the time series data and make predictions. Overall, the research paradigm involves the collection and analysis of quantitative data using statistical and econometric techniques to predict private consumption.

3.3 Data Collection

This section presents the relevant variables and data used to build the predictive model. They consist of online search data, conventional economic data, the consumer confidence sentiments index, and macroeconomic variables. The following subsections provide detailed information on which data to collect and how.

3.3.1 Online Search Data

In compiling the online search data for this study, various sources such as Mastercard transactions are considered. Considering the limited time and resources as well as data availability, it was decided to use Google search data as online search data. This is the volume of data that users' access through Google on a daily basis.

In Malaysia, Google search is the predominant online search platform (Comscore, 2023; Semrush, 2023), with a large majority of users using Google Chrome as their preferred browser (StatCounter, 2023). Given the widespread use of Google search, it serves as a valuable source of information on consumer behaviour and preferences. By focusing on Google search data, this study aims to provide insights into search behaviour related to

various daily needs such as food, clothing, housing, transportation, education, and entertainment. This study recognises the importance of the Google search engine as the primary observation platform for the creation of search engine indicators.

Among them, the connotation of constructing search questions related to private consumption as indicators. Vosen and Schmidt (2011) pointed out that search attention can be used as a pre-purchase action for consumers, therefore, in terms of the measures constructed for this study to obtain Google search data from Google Trends, the next challenge is which keywords used to search in Google Trends to download the relevant data.

When selecting search terms for Google Trends, the search terms that are close to consumer behaviour are considered. Therefore, the top e-commerce website was published in the "Top eCommerce & Shopping Websites in Malaysia Ranking Analysis for July 2023" by Similarweb (2023). This study selected the most popular website – Shopee, with the total of 41.04 million monthly visits, with an average of 9.12 pages viewed and an average dwell time of almost 9:18 minutes per visit.

Following Matsumoto et al.'s (2013) method of using product category names as keywords, the Shopee website, the most visited marketplace in Malaysia, is used as the source for keywords in the selection process. The names of all categories must be consistent with COICOP (United Nations, 2018), which is also used by the Department of Statistics Malaysia.

The product categories on the selected website are divided into main categories, subcategories, brand lists, and conspicuous products; Bahasa Malaysia language was translated and combined with English category. When analysing the research results, it is

important to remember that too many keywords and data collection will affect the analysis result. Therefore, it was decided to use the middle product category or subcategory as the basis for selecting keywords in Google Trends searches. The Shopee website has a total of 24 main categories and 248 subcategories. These unique subcategories are used as keywords for Google Trends.

3.3.2 Conventional Economic Data

Considering the selection of conventional economic indicators to effectively exercise the effect of consumption forecasting, based on the Bank Negara Malaysia (BNM) report, there are percentage annual changes in sales of passenger vehicles, consumption credit, imports of consumption goods, and outstanding balance of credit cards recorded in the BNM statistics in the private consumption section. In addition to the various indicators in the private consumption section of the BNM report used for this study, the producer price index, total industrial production index, trade exports, trade imports, unemployment, and labour force are also used for this study to show market performance in various areas. As shown in Table 3.1, all the above variables are listed with the source of the data obtained and all data frequencies are a mixture of monthly and quarterly.

Table 3.1: Data Construction for Conventional Economic Indicator

No	Items	Code	Frequency
1	Sales of Passenger Cars (annual change %)	CEcar	Monthly

2	Consumption Credit (annual change %)	CEccredit	Monthly
3	Import of Consumption Goods (annual change %)	CEimport	Monthly
4	Outstanding Balance of Credit Card (annual change %)	CEowing	Monthly
5	Producer Price Index (annual change %)	CEppi	Monthly
6	Industrial Production Index (Index 2015=100, annual change %)	CEipi	Monthly
7	Trade Exports (annual change %)	CEtexport	Monthly
8	Trade Imports (annual change %)	CEtimport	Monthly
9	Unemployment (annual change %)	CEunemploy	Quarterly
10	Labour Force (‘000 persons, annual change %)	CElabour	Quarterly

Notes: Items 1 to 4, reference from Choi and Varian (2012); data source from Section 3.5.6: Private Consumption Indicators, Monthly Highlights and Statistics of Bank Negara Malaysia. Available at <https://www.bnm.gov.my/-/monthly-highlights-statistics-in-november-2023>

Items 5 to 10, reference from Choi and Varian (2012), Pavlicek (2014); data source from Section 3.5: Selected Economic Indicators, Monthly Highlights and Statistics of Bank Negara Malaysia. Available at <https://www.bnm.gov.my/-/monthly-highlights-statistics-in-november-2023>

3.3.3 Consumer Confidence Sentiments Index

The Consumer Confidence Sentiment Index serves as a valuable source of information on the current level of optimism among consumers. It indicates whether the general mood and willingness of consumers to spend money is improving or deteriorating over time. The main objective of this index is to gain insights into and monitor consumers' behaviour and sentiment. It should be noted that the Consumer Confidence Sentiment Index is subjective and reflects market confidence rather than actual market performance. Consequently, it does not fall within the scope of conventional economic data.

In the context of predicting consumption patterns, Vosen and Schmidt (2011) and Matsumoto et al. (2013) recommend using the consumer confidence sentiment index as an observable variable. The Malaysian Institute of Economic Research provides the Consumer Confidence Sentiment Index on a quarterly basis, which ranges from 0 to 200 and has a neutral point of 100. In order to standardise the data for all variables, a conversion from quarterly to monthly frequency is required.

3.3.4 Macroeconomic Variables

This study includes both private consumption expenditure as the response variable and gross domestic product (GDP) as the explanatory variable, as shown in Table 3.2.

Table 3.2; both data are from Section 3.3, Gross Domestic Product by Expenditure Component at Constant 2015 Prices (Annual Change) of BNM's Monthly Highlights and

Statistics. Both data are reported on a quarterly basis, and in order to standardise the data for all variables, a conversion from quarterly to monthly frequency is required.

Table 3.2: Data Construction for Macroeconomic Variables

No	Items	Code	Frequency
1	Private Final Consumption Expenditure Indicator (Index 2015=100, annual change %)	PCI	Quarterly
2	Gross Domestic Product (Index 2015=100, annual change %)	GDP	Quarterly

Notes: Items 1 and 2, reference from Kholodilin, Podstawski, and Siliverstovs (2010); data source from Section 3.3: Gross Domestic Product by Expenditure Components at Constant 2015 Prices (Annual Change), Monthly Highlights and Statistics of Bank Negara Malaysia. Available at <https://www.bnm.gov.my/-/monthly-highlights-statistics-in-november-2023>

3.3.5 Sample Period

According to time series experts, a minimum of 50 observations is recommended when conducting time series analysis (Warner, 1998). These guidelines are particularly important when seasonal components are involved, as monthly observations over several years are required to accurately capture and model such patterns.

In the case of this study, there are limitations to the availability of the data. The GDP data starts on 1 January 2016, after the rebasing to 2015 was completed. Similarly, improvements were made to the Google data collection system from 1 January 2016. Therefore, all data used in this study extend from 1 January 2016 to 31 March 2023, resulting in a total of 87 observations. Although this is above the minimum requirement of 50

observations for time series analysis, it is important to note that the data frequency is in monthly form.

3.4 Data Analysis

PCA or principal component analysis is a statistical technique that can be used to extract the underlying factors from a large number of Google Trends keywords related to private consumption. It involves transforming the original set of variables into a new set of variables called principal components, which are linear combinations of the original variables. The first principal component explains the largest variation in the data, the second component explains the largest variation among the remaining variables, and so on. Using the PCA, we can identify the most important keywords related to private consumption that are highly correlated with each other and group them into a smaller number of factors that capture the most important patterns of variation in the data.

In this study, PCA is used to extract the large components of the time series data from both the survey-based indicators and the Google Trends data. These extracted components can then be used as predictors in a time series forecasting model to predict private consumption. Using the PCA, the model can capture the underlying patterns and trends in the data, resulting in more accurate forecasts.

3.4.1 Static Extraction of Online Search Indicators

In this study, product category names from the Shopee website are used as search terms entered into Google Trends to download relevant data; details are explained in Section 3.3.1. After using the statistical tool, PCA is performed to extract the components that constitute the online search indicator; the number of factors extracted is determined according to the eigenvalue and the factor's ability to explain the overall variation. The screening principle is also to select factors with an eigenvalue greater than 1 or it means that the cumulative percentage of total variance explained is greater than 70%.

3.4.2 Static Extraction of Conventional Economic Indicator

In addition to the established online search indicators mentioned above, this study collects leading economic data as conventional economic indicators for forecasting private consumption (explained in detail in Section 3.3.2). First, PCA of the conventional economic indicators is conducted. PCA reduces the variables from complex to simple variables, as well as concentrates and refines the data of the conventional economic indicators.

In extracting the conventional economic indicators, the consumer confidence sentiment index (explained in Section 3.3.3) is considered to show consumers' current views of the market, which is a subjective measure, the desired stage in the AIDA model. Various measures in the conventional economic indicators listed in Section 3.3.2 are taken from both the section on private consumption indicators and the section on selected economic indicators in BNM's Monthly Highlights and Statistics publication. Therefore, in the

selection of components, the consumer confidence index is not considered as a measure of the conventional economic indicators. Based on characteristics of the value (eigenvalue) and the ability of the factor to explain the total variation, the number of factor extractions is determined. The screening principle is to select factors with an eigenvalue greater than 1 or a cumulative contribution greater than 70%.

3.4.3 Online and Conventional Private Consumption Models

This section begins with a comprehensive analysis based on a number of data sources and indicators, as shown in Table 3.3. These sources include online search indicators (as described in Section 3.4.1), conventional economic indicators (as described in Section 3.4.2), the consumer confidence sentiment index (as described in Section 3.3.3) and macroeconomic variables (as described in Section 3.3.4).

Table 3.3: Summary of Variables Used for Time Series Analysis

Code	Name	Source	Reference
OS1	Online searches indicator related to personal and household goods	PCA extracted from this study	Matsumoto, Matsumura and Shiraki (2013)
OS2	Online searches indicator related to transport	PCA extracted from this study	Matsumoto, Matsumura and Shiraki (2013)
CE1	Conventional economic indicator reflects the volume and value of goods and services	PCA extracted from this study	Matsumoto, Matsumura and Shiraki (2013)
CE2	Conventional economic indicator reflects the production and demand of goods and services	PCA extracted from this study	Matsumoto, Matsumura and Shiraki (2013)

CCSI	Consumer Confidence Sentiment Index	MIER	Juhro and Lyke (2020)
GDP	Gross Domestic Product	DOSM	Kholodilin, Podstawski and Siliverstovs (2010))
PCI	Private Consumption Indicator	DOSM	Kholodilin, Podstawski and Siliverstovs (2010)

This multi-layered analysis aims to construct two different models, namely the online private consumption model and the conventional private consumption model to capture the complex dynamics between the different time series variables, thus provide a basis for correct analysis and predictions.

To gain a comprehensive understanding of these interrelated variables, a multi-stage analytical approach is used for both models. In these stages, the complex interactions, interdependencies, causal relationships and long-term equilibrium patterns that characterise these time series variables are explored to decipher the underlying dynamics and provide valuable insights that enable robust analysis and prediction.

3.4.3.1 Unit Root Test

The analysis steps include a critical unit root test, which is a fundamental component of time series analysis. This test serves the crucial purpose of determining whether a particular time series variable is stationary or non-stationary. Stationarity plays a central role in time series analysis because it affects the statistical properties of a variable over time. If a variable is non-stationary and has a unit root, it means that its behaviour changes over time and it is dependent on past values. Importantly, a non-stationary regressor can invalidate

many standard empirical results (Dritsakis, 2004). Therefore, a unit root test is essential to confirm the suitability of the data for analysis.

In this study, we use three well-established unit root tests: the Augmented Dickey-Fuller (ADF) test introduced by Dickey and Fuller (1979), and the Dickey-Fuller Generalised Least Squares (DFGLS) test proposed by Elliott et al. (1996). These tests are valuable tools for assessing stationarity characteristics.

Both the ADF and DFGLS tests have a common null hypothesis revolving around the presence of a unit root. However, they differ in how they specify the alternative hypothesis and how they deal with deterministic components in the data generation process (DGP) of a variable. In particular, the DFGLS method involves local detrending and/or detrending of a series before the usual ADF auxiliary regression is performed. This approach proves advantageous when examining series that have a mean and/or a linear trend in their DGP as it increases statistical power and minimises the risk of false conclusions (Elliott et al., 1996).

3.4.3.2 Cointegration Test

If a unit root test reveals the non-stationarity of several variables, this does not necessarily rule out the existence of a significant long-run equilibrium relationship between them. Cointegration is a key statistical concept that sheds light on such scenarios. It states that two or more time series, typically of order 1 (denoted $I(1)$), can be non-stationary but still exhibit a stable and significant long-run equilibrium relationship. This property is particularly valuable for understanding the dynamics of economic and financial systems.

In this study, the system-based cointegration procedure introduced by Johansen and Juselius (1990) is used to rigorously investigate the presence or absence of long-run equilibrium. This approach offers several advantages over alternative methods. In particular, it does not depend on the choice of normalisation, which can be a source of ambiguity in other approaches. It is also more robust than the Engle-Granger test introduced by Gonzalo (1994). Phillip's (1991) study, emphasised the attractiveness of this technique because of its properties of symmetry, unbiasedness, and efficiency.

The core of this cointegration test procedure lies in the use of two critical Likelihood Ratio Test statistics (LR), namely the Trace Test and the Maximum Eigenvalue Test. These statistics serve as reliable tools for evaluating the number of cointegrating vectors present in the system. By examining the Trace Test and the Maximum Eigenvalue Test, the extent of the long-term equilibrium relationships between the variables under study can be effectively determined, providing valuable insights into their interdependencies and dynamics.

3.4.3.3 Granger Causality Test

The causality test plays a crucial role in uncovering the intricate web of causal relationships that exist between the variables. Its primary goal is to determine whether changes in one variable are due to changes in another and effective in examining the cause-and-effect dynamics within a system. A widely used framework for causality testing is Granger causality, which is a concept based on the idea that a causal relationship can exist between variables if past values of one variable can help predict future values of another variable.

However, when cointegration is established, the application of Granger causality becomes a nuanced endeavour that requires its integration into the Vector Error Correction Model (VECM). This step is crucial to circumvent potential problems of model misspecification, as Granger (1988) pointed out in his seminal paper. The VECM, a special variant of the vector autoregressive model (VAR), contains the crucial feature of cointegration between variables. To ensure the accuracy and relevance of the model, it is essential to include the relevant error correction terms (ECTs) in the VAR framework and thus avoid omitting important constraints.

VECM is a versatile analytical tool that can be used to investigate both short-term dynamics by including lagged differences of variables and long-term equilibrium relationships by examining cointegrating vectors. In the context of a cointegration model, the application of Granger causality in the VECM framework is essential to assess the significance of ECTs. This approach is particularly important because the presence of a cointegrating relationship in the long run means that the residuals derived from the cointegrating equation can effectively act as error correction terms that can improve the model's ability in capturing and interpreting causal relationships. VECM tests provide further insight into the dynamics of the cointegrated system and how the variables influence each other over time.

3.4.3.4 Variance Decomposition for Future Analysis

The VECM is a valuable tool for demonstrating Granger causality within a given sample period. However, it has certain limitations in that it is unable to shed light on the

relative contributions of the explanatory variables in explaining variations in the dependent variable beyond this limited period, as noted by Masih et al. (1997). To address this limitation, we have incorporated variance decomposition diagrams (VDCs) developed by Sims (1980). With these techniques, we can not only detect the presence of causal relationships in the Granger causal chain, but also measure the degree of exogeneity between variables beyond the sample period, according to the methodology described by Masih and Masih (2001).

The use of forecast error VDCs allows us to draw inferences about the proportion of movement in a time series that is due to its inherent shocks, as opposed to shocks that are due to external influences within the system, as described by Enders (2015). These VDCs provide an innovative means of presenting this information in percentage form. The strength of the relationships between the five variables, both in terms of their own shocks and their interactions with each other, is quantified on a scale from 0% to 100%. For a variable that is optimally forecast using its own lagged values, the entire variance of the forecast error is explained by its own disturbances, as originally postulated by Sims (1982).

In this analysis, we use VDCs over different time horizons, ranging from 1 to 50 months, to comprehensively assess the dynamic interaction and explanatory power of the variables within the system.

3.5 Research Framework

Private consumption forecasting provides MSMEs with valuable insights into consumer behaviour that enable them to make informed decisions about sales targets, budget allocations, and pricing strategies. It enables MSMEs to optimise their resources by matching production capacity, labour, and inventory levels to expected consumer demand, thus minimising unnecessary expenditure and maximising operational efficiency. In addition, private consumption forecasting helps MSMEs identify untapped markets, niche segments and changing consumer needs, enabling them to develop innovative products or services and take advantage of market opportunities. By providing timely and relevant offerings, MSMEs can effectively meet customer expectations and secure a competitive advantage in the dynamic business landscape.

The AIDA model states that once consumers become aware, they are interested in learning more. Searching for information on goods or services is the service activity that accounts for 92.5 per cent of the ICT Access and Use Report (Department of Statistics Malaysia, 2023, May 31); Individuals and Households Survey. Online search data can be used to gain insights into the factors that influence private consumption. By analysing the frequency of specific search terms related to products or services, researchers can identify trends and patterns that may indicate changes in consumer preferences. For example, an increase in searches for "home exercise equipment" may indicate a growing interest in home fitness solutions, which in turn could influence private consumption patterns in the fitness industry.

Figure 3.1 is the research framework produced for this study. As described in Section 3.3.1, the online search data is the frequency of keywords searched by Malaysians in Google search. The keywords downloaded from Google Trends are the names of sub-categories of products and services listed on the Shopee website. All keywords selected and used fall within the scope of COICOP (United Nations, 2018), which are closer to people's consumption compared to the names of all categories from COICOP 2018. After downloading all the data, Section 3.4.1 describes the use of PCA to reduce, focus, and refine the data from complex to simple, with the extracted components constituting the online search indicator for further analysis.

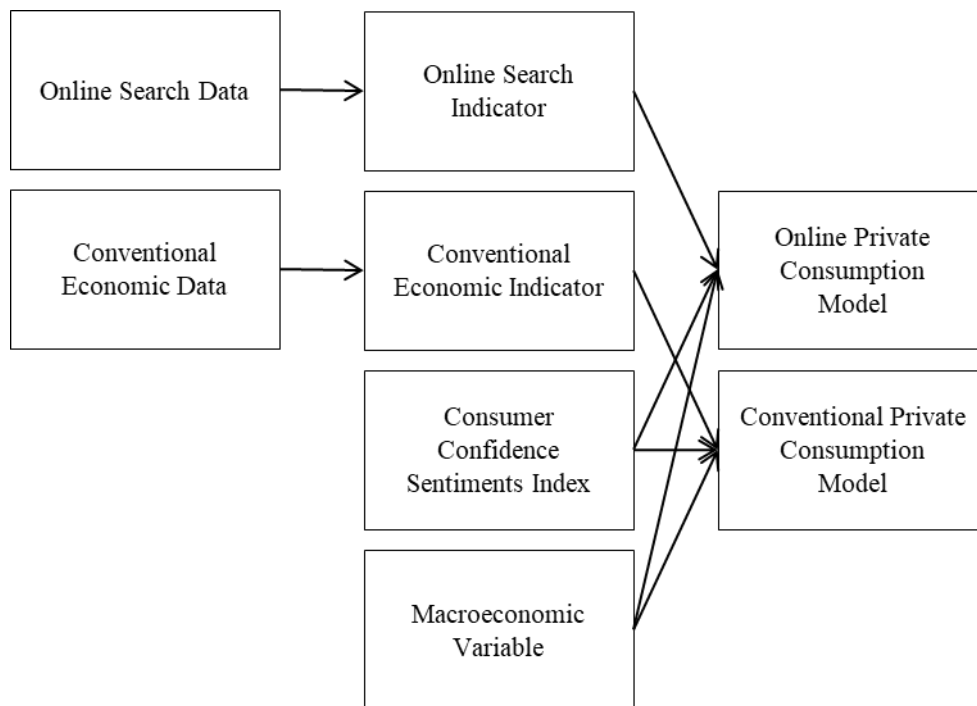


Figure 3.1: Research Framework

Conventional economic indicator treated as the action, the final 'A' in AIDA model. as described in Section 3.3.2 is to acquire leading economic measures from DOSM and BNM. Section 3.4.2 uses PCA to extract components representing conventional economic indicators.

Consumer confidence sentiment index is considered the “D” or the desire in the AIDA model. It is about people’s mood, and their subjective view of the current market sentiment. As described in Section 3.3.3, secondary data must be obtained from CEIC Data and MIER, both of which must be converted from quarterly to monthly data. The macroeconomic variables are private consumption indicators and selected economic indicators from DOSM and BNM, as described in Section 3.3.4.

After the static extraction of the online search indicators and the conventional economic indicators, the analysis of the predictive model for the online and conventional predictive indicators is conducted to discuss the improvement of the predictive model of private consumption and to evaluate whether the application of online search data can support the prediction of private consumption in Malaysia. The empirical analysis is based on the evaluation of the combination of online search indicators (Section 3.4.1), conventional economic indicators (Section 3.4.2), online search indicators (Section 3.4.1), consumer confidence sentiment index (Section 3.3.3) together with macroeconomic variables (Section 3.3.4); the estimation using VAR model analysis to discuss the predictive effect of the online and conventional model of private consumption.

3.6 Summary

In conclusion, this chapter has provided a comprehensive overview of the methodology used in this study to forecast private consumption in Malaysia. The research paradigm, design, and data collection processes have been outlined, focussing on the integration of online search data, conventional economic indicators, consumer confidence sentiment index, and macroeconomic variables.

The research adopted a quantitative and empirical approach within a positivist epistemological framework. Time series analysis, in particular principal component analysis (PCA), formed the backbone of the analytical methods used. The chapter describes the selection and extraction of data from various sources, including Google Trends for online search data, conventional economic indicators from Bank Negara Malaysia (BNM), and consumer confidence sentiment data from the Malaysian Institute of Economic Research (MIER).

The detailed discussion about the construction of the search queries with Google Trends, the selection of relevant keywords and the reasons for the selection ensured transparency in the data collection process. The inclusion of PCA for both the online search indicators and the conventional economic indicators aimed to reduce complexity and identify key components for further analysis.

The methodological framework also included critical steps such as unit root tests, cointegration tests, Granger causality tests, and variance decomposition for a thorough examination of relationships and predictive ability. The research design involved the creation of two predictive models, namely the online model of private consumption and the

conventional model of private consumption, which both capture the complex dynamics between different time series variables.

The chapter concludes by presenting a robust research framework based on the AIDA model, which illustrates the role of online search data, conventional economic indicators, consumer confidence sentiment, and macroeconomic variables in predicting private consumption. The framework serves as a guide for the subsequent empirical analysis and facilitates the evaluation of the models.

CHAPTER 4

DATA ANALYSIS AND FINDINGS

4.1 Overview

In this study, the indicator for predicting private consumption is supported by indicators of online search in addition to conventional economic indicators. In order to incorporate online search into an effective indicator for predicting private consumption, this study uses PCA to reduce the variables from complex to simple and to condense and refine the information of the variables for operational analysis. In the first section, the online search and conventional economic indicators are extracted using PCA. In the following section, both the online private consumption model and the conventional model are subjected to various tests to compare both models and their predictive power for private consumption is discussed.

4.2 PCA Extraction of Online Search Indicator

4.2.1 Data Extraction from Google Trends

After acquiring all subcategories of products listed in Shopee website, it is treated as keywords to search in Google Trends, allowing the researcher to download trending data for each specific keyword. The data is related to the popularity or interest of that keyword over time, within the Malaysia region, and in related Google Trends categories, monthly search volume data from January 2016 to March 2023. 87 sample size was obtained for all data.

While PCA can be conducted with a sample that is less than 100 cases but more than 50 cases, caution must be taken in interpretation. A total of 248 monthly time series data were downloaded, categorised by COICOP division.

Each division in the COICOP classification corresponds to a particular category of consumption item. Table 4.1 is a summary of the divisions in the COICOP classification along with the number of keywords extracted with each division. The keywords were extracted from the subcategories of the products on Shopee.com.my and grouped based on the COICOP division, with a larger number of keywords such as clothing and footwear divided into two categories each and fewer keywords of 1 or 2 not processed further with PCA.

Table 4.1: Number of Keywords for Online Search Data

COICOP Division	COICOP Description	No. of Keywords	PCA?
1	Food and Beverages	15	Proceed
2	Alcoholic and Tobacco	1	
3a	Clothing	50	Proceed
3b	Footwear	34	Proceed
4	Housing, Utilities	6	Proceed
5	Household Equipment, etc.	51	Proceed
6	Health	22	Proceed
7	Transport	13	Proceed
8	Info. and Comm.	13	Proceed
9	Recreation, Sport	41	Proceed
10	Education Services	2	
Total Keywords		248	

4.2.2 Results from Each Products Categories

After PCA was performed for each of the categories, a total of 9 PCAs were performed and the factor scores loaded in the first component were calculated in each individual category. Then, all the results are summarised. First, the adequacy of the sample for the online search data is examined. This is the adequacy of factor analysis, a measure of how well the online search data can be used for factor analysis, a statistical technique used to identify the underlying dimensions or factors that explain the correlations between a set of variables.

One way to assess sample adequacy for online search data is to use the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test for sphericity. The KMO is a statistic that indicates the proportion of variance in variables that could be caused by underlying factors. High values, which are values closer to 1, generally indicate a better fit of the sample for factor analysis. If the value is below 0.50, the results of the factor analysis are probably not very meaningful. Bartlett's test for sphericity tests the hypothesis that your correlation matrix is an identity matrix, which would indicate that your variables are not related and therefore not suitable for factor analysis. Small values (less than 0.05) of the significance level indicate that the data are suitable for factor analysis.

Table 4.2 presents the results of the KMO and Bartlett's tests as measures of sample adequacy for online search data for each division of the COICOP classification. These measures help in assessing the suitability of the data for conducting PCA.

Table 4.2: Sample Adequacy for Online Search Data

COICOP Division	COICOP Description	KMO	Bartlett
1	Food and Beverages	0.580	0.000
3a	Clothing	0.812	0.000
3b	Footwear	0.704	0.000
4	Housing, Utilities	0.591	0.027
5	Household Equipment, etc.	0.663	0.000
6	Health	0.652	0.000
7	Transport	0.784	0.000
8	Info. and Comm.	0.610	0.000
9	Recreation, Sport	0.589	0.000

Table 4.2 shows that all COICOP divisions have KMO values above 0.50 and Bartlett's test significance levels below 0.05, suggesting that the online search data are highly correlated and that factor analysis may be appropriate for these divisions. However, some divisions have higher KMO values than others, suggesting that they have more appropriate data for factor analysis. For example, Clothing has the highest KMO value of 0.812, meaning that a high proportion of the variance is explained by the underlying factors, while Leisure and Sport has the lowest KMO value of 0.589, which means a lower proportion of the variance is explained by the underlying factors.

Online search data is a type of data obtained from Google search queries and it also presents some challenges for reliability and validity. Reliability and validity are important concepts in quantitative research (SÜRÜCÜ & MaslakÇI, 2020), since both indicate the effectiveness of a method, procedure, or test measures. Reliability is about the consistency of a measurement while validity is about the accuracy of a measurement.

In the given Table 4.3, the reliability and validity of the online search data for each division in the COICOP classification are assessed based on the number of items contributing to the first principal component by PCA, together with Cronbach's alpha, the maximum, and minimum value of the factor value in the first component to detect outliers, and whether the division was included or selected in the next PCA of the online search indicator.

Table 4.3: Reliability and Validity Measures for Online Search Data

COICOP Division	COICOP Description	Items loaded in 1st Component	Cronbach's Alpha	Max. Score	Min. Score	PCA?
1	Food and Beverages	5	0.759	2.157	-2.497	Included
3a	Clothing	16	0.938	3.242	-1.492	Included
3b	Footwear	10	0.862	2.578	-1.786	Included
4	Housing, Utilities	4	0.193	4.651	-1.385	
5	Household Equipment, etc.	9	0.895	2.340	-1.829	Included
6	Health	4	0.803	2.120	-1.569	Included
7	Transport	8	0.866	2.325	-2.344	Included
8	Info. and Comm.	3	0.697	2.962	-1.808	Included
9	Recreation, Sport	6	0.759	2.435	-1.687	Included

Cronbach's alpha is a measure of the internal consistency or reliability of the data. A Cronbach's alpha of 0.60 or higher is the minimum acceptable level. An alpha of 0.70 or higher is preferred. One way to detect outliers is to check the values for each factor that are greater than 3.0 or less than -3.0, which is outside the acceptable range. As shown in Table 4.3, apart from the Housing Utilities variable whose alpha is 0.193, all the other variables are above 0.60 of the minimum acceptable level; more than half of the variables have relatively high Cronbach's alpha values of above 0.80. The Housing Utilities variable also

shows that one of its factor values is outside the acceptable range of 3.0 and below, which may lead to lower reliability and validity. Therefore, this variable was excluded for further analysis. Clothing is another variable where the factor value of 1 period is above 3.0, where the measure of sampling adequacy for the Kaiser-Meyer-Olkin test is 0.812, which exceeds the minimum requirement of 0.5 for the entire set of variables; and the probability associated with Bartlett's test of sphericity is less than 0.001, which meets the adequacy of factor analysis, so this variable is included for further analysis. The remaining factor values of the variables are in the acceptable range, indicating good internal consistency or reliability of the data. Therefore, except for the variables 'housing' and 'care', all other variables are included in the PCA for the online search indicators.

4.2.3 Computation for Online Search Indicator

Based on the results of the first principal component of each COICOP division, the following 8 variables are factorised from the PCA of each COICOP division. After passing the validity and reliability tests, they are included in the next PCA of indicators for online search. These 8 variables are “food and beverages” [OSM1], “clothing” [OSM3A], “footwear” [OSM3B], “household appliances, etc.” [OSM5], “Health” [OSM6], “Transport” [OSM7], “Info and Comm.” [OSM8], and “Recreation, sports” [OSM9]. The sample size in each case remains at 87. The ratio of cases to variables in a PCA should be at least 5 to 1. With 87 and 8 variables, the ratio of cases to variables is 10.9 to 1, which exceeds the requirement for the ratio of cases to variables. In iteration 1, the communality for the variable

"Info. and Comm." [OSM8] was 0.359, which is less than 0.50. Therefore, the variable was removed and the PCA was calculated again.

Factor analysis adequacy (PCA) requires that the Kaiser-Meyer-Olkin measure of sampling adequacy is greater than 0.50 for each individual variable as well as for the set of variables. The measure of sampling adequacy is described as marvellous if it is 0.90 or more, meritorious if it is in the range of 0.80, middling if in the 0.70's, mediocre if in the range of 0.60, miserable if it is in the range of 0.50, and unacceptable if it is below 0.50. The overall measure of sampling adequacy for the variables included in the analysis was 0.861, which is above the minimum requirement of 0.50 for the overall measure of sampling adequacy. In addition, PCA requires that the probability associated with Bartlett's test of sphericity is below the significance level. The probability associated with Bartlett's test is < 0.001 , which fulfils this requirement. Thus, the seven variables remaining in the analysis meet the criteria for adequacy of factor analysis. The next step is to determine the number of factors to be included in the factor solution.

In the analysis, there were no variables with communalities of less than 0.5. The components explain at least 50% of the variance in each of the variables. In calculating the PCA, the number of components extracted is determined according to the eigenvalue and the ability of the component to explain the total variation. The screening principle is also to select components with an eigenvalue greater than 1, or it means that the cumulative percentage of total variance explained is greater than 70%.

The explained total variance for the online search indicator is the breakdown of the variance captured by each PCA. It shows the contribution of each component to the total

variance in the data. Table 4.4 shows the most important components, eigenvalues, percentage of variance and cumulative percentage for interpreting the results.

Table 4.4: Total Variance Explained for Online Search Indicator

Component	Eigenvalues	% of Variance	Cumulative %
1	4.764	68.060	68.060
2	1.050	15.000	83.060
3	0.393	5.607	88.667
4	0.316	4.515	93.181
5	0.239	3.410	96.591
6	0.166	2.374	98.965
7	0.072	1.035	100.000

The components column shows the number of components extracted from the data. The eigenvalues column indicates the proportion of variance captured by each principal component. The eigenvalues are proportional to the sum of the squared charges of each component. The larger the eigenvalue, the more important the component. The eigenvalue greater than 1 is used as a screening principle, which is a common criterion for selecting the number of components (Sallis et al., 2021). There are seven components, since the eigenvalue is less than 1 from the third component, two components are extracted. In this case, the first component has an eigenvalue of 4.764, which means that it explains 68.060% of the total variance. The second component has an eigenvalue of 1.050, which means that it explains 15.000 % of the total variance, and so on.

The percentage of variance column shows the percentage of each component in the total variance, taking into account the percentage of variance and the cumulative value. It is

determined by dividing the eigenvalue of each component by the sum of all eigenvalues. In this case, the sum of all eigenvalues is 7, so the percentage of variance explained by each component is equal to its eigenvalue divided by 7. For example, the percentage of variance explained by the first component is $4.764 / 7 = 68.060\%$.

The cumulative percentage column shows the cumulative percentage of the total variance explained up to that particular component. It indicates how much of the total variance is explained by a given number of factors. In this case, the cumulative percentage of variance explained by the first two factors is 83.060%, which means that these two factors capture most of the variability in the data. As we move down the list, each successive component explains a decreasing proportion of the variance. At the seventh component, the cumulative percentage reaches 100%, which means that all the variance in the data was explained by these seven components.

Table 4.5 shows the factor loadings for an indicator of online search, which indicates how often people search for a particular topic on the Internet. Factor loadings are the correlations between the observed variables and the latent factors in the PCA. They indicate how strongly each variable is associated with each factor. The higher the absolute value of the factor loading, the more important the variable is for the definition of the factor.

The information in 7 of the variables can be represented by 2 components and 1 single variable. Component 1 seems to represent online searches related to personal and household goods. It has high positive loadings on variables such as “Household Equipment, etc.” [OSM5], “Footwear” [OSM3B], “Clothing” [OSM3A], “Health” [OSM6] and “Recreation, Sport” [OSM9]. These are all topics that people might search for when they

want to buy or find out about products or services that affect their personal well-being or lifestyle. In further analyses, this combination of variables can be replaced by a component variable [OS1].

Table 4.5: Factor Loadings for Online Search Indicator

Code	Description	Component 1 (OS1)	Component 2 (OS2)
OSM5	Household Equipment, etc.	0.909	
OSM3B	Footwear	0.903	
OSM3A	Clothing	0.869	
OSM6	Health	0.779	
OSM9	Recreation, Sport	0.672	
OSM7	Transport		0.957
OSM1	Food and Beverages		0.679

Note: Based on PCA extraction method, and Varimax with Kaiser Normalisation rotation method

Component 2 seems to represent online searches related to transport. It has high positive loadings on variables such as “Transport” [OSM7] and “Food and Beverages” [OSM1]. These are themes that people might search for when they want to travel or eat out. In further analyses, this combination of variables can be replaced by a component variable [OS2]. The variables “Info. and Comm.” [OSM8] was not included in the components and is retained as a single variable. The components explain 83.06% of the total variance of the variables included in the components.

4.3 PCA Extraction of Conventional Economic Indicator

6 of the variables recorded for the PCA are “Outstanding Balance of Credit Card” [CEowing], “Labour Force” [CElabour], “Unemployment” [CEunemploy], “Consumption Credit” [CEccredit], “Trade Imports” [CEimport], “Trade Exports” [CEtexport], “Producer Price Index” [CEppi], “Import of Consumption Goods” [CEimport], “Sales of Passenger Cars” [CEcar], and “Industrial Production Index” [CEipi]; monthly period from January 2016 to March 2023, the number of samples is 87. The data exceed the requirement for the ratio of cases to variables of 5 to 1.

For a total of 4 iterations, the communality for the variables “Outstanding Balance of Credit Card” [CEowing], “Labour force” [CElabour], “Unemployment” [CEunemploy], and “Consumption Credit” [CEccredit] was less than 0.50. Therefore, the variables were removed and the PCA was recalculated.

The overall measure of sampling adequacy for the variables included in the analysis was 0.734, which is above the minimum requirement of 0.50 for the overall measure of sampling adequacy. Furthermore, the probability associated with Bartlett's test < 0.001 , meeting the requirement. Thus, the six variables remaining in the analysis meet the criteria for factor analysis adequacy. The next step is to determine the number of factors to be included in the factor solution.

The components explain at least 50% of the variance in each of the variables. Table 4.6 shows the variance explained by each component for the conventional economic indicator used to assess, measure and evaluate the overall health of the economy. It indicates how much of the total variance in the data is accounted for by each component. Each row

represents a different component (from 1 to 6) and the corresponding columns provide information on the eigenvalue, the percentage of variance explained by that component and the cumulative percentage of variance up to that component.

Table 4.6: Total Variance Explained for Conventional Economic Indicator

Component	Eigenvalues	% of Variance	Cumulative %
1	3.747	62.446	62.446
2	1.202	20.027	82.473
3	0.617	10.279	92.752
4	0.249	4.154	96.906
5	0.117	1.957	98.863
6	0.068	1.137	100.000

Table 4.6 shows that Component 1 has an eigenvalue of 3.747, which accounts for 62.446% of the total variance in the data. Component 2 adds another 20.027% of variance to the previous component, resulting in a cumulative percentage of 82.473%. This process continues for each subsequent component until component 6, which accounts for 1.137% of the total variance, bringing the cumulative percentage to 100.000%. The cumulative percentage shows that the first two components account for 82.473% of the total variance.

Table 4.7 shows the factor loadings for a conventional economic indicator on two components, which are Component 1 and Component 2. Factor loadings of the correlation between each variable and the underlying factors.

Table 4.7: Factor Loadings for Conventional Economic Indicator

Code	Description	Component 1 (CE1)	Component 2 (CE2)
CEimport	Trade Imports	0.970	
CEtexport	Trade Exports	0.866	
CEppi	Producer Price Index	0.856	
CEimport	Import of Consumption Goods	0.726	
CEcar	Sales of Passenger Cars		0.946
CEipi	Industrial Production Index		0.800

Note: Based on PCA extraction method, and Varimax with Kaiser Normalisation rotation method

Component 1 includes the variables “Trade Imports” [CEimport], “Trade Exports” [CEtexport], “Producer Price Index” [CEppi] and “Import of Consumption Goods” [CEimport]. These are all indicators that reflect the volume and value of goods and services exchanged between trading partners. A higher value of this factor means that the economy is more open and competitive in the market. In further analyses, this combination of variables can be replaced by a component variable [CE1].

Component 2 comprises the variables “Sales of Passenger Cars” [CEcar] and “Industrial Production Index” [CEipi]. seems to represent production. It has high positive loadings on variables such as Sales of Passenger Cars and Industrial Production Index. These are indicators that reflect the production and demand of goods and services within the country. A higher value of this factor means that the economy is more productive and prosperous. In further analyses, this combination of variables can be replaced by a component variable [CE2].

The variables “Outstanding Balance of Credit Card” [CEowing], “Labour Force” [CElabour], “Unemployment” [CEunemploy] and “Consumption Credit” [CEccredit] were not included in the components and are kept as single variables. The components explain 82.47% of the total variance of the variables included in the components.

4.4 VAR Online and Conventional Model Analysis

4.4.1 Unit Root Test

A unit root test is a statistical procedure for assessing the non-stationarity of a time series, which may take the form of changes in mean, variance, or autocorrelation over time. The presence of a unit root makes a time series unpredictable and unreliable for forecasting. Unit root tests, including the Augmented Dickey-Fuller (ADF) test, and the Dickey-Fuller Generalised Least Squares (DFGLS) test, serve as tools for determining whether a time series variable is stationary or has a unit root. The ADF test uses an autoregressive model to determine whether a unit root exists and tests the null hypothesis of a unit root against the alternatives of stationarity or trend stationarity. Meanwhile, the DFGLS test, a further development of the ADF test, increases its explanatory power by applying a detrending procedure according to the general least squares (GLS) method before testing for the presence of a unit root.

Unit root tests usually serve as the first step in the modelling process. If a variable is found to have a unit root, appropriate transformations or differentiation methods can be applied to make it stationary and thus suitable for more robust analysis and forecasting.

Table 4.8 provides a comprehensive overview of the results of various unit root and stationarity tests applied to different time series variables in two different scenarios: A) Levels and B) First Differences. These tests rely on six key statistics: τ_{μ} , τ_{τ} , τ_{μ} , τ_{τ} , η_{μ} and η_{τ} , which correspond to the ADF and DFGLS test statistics designed to detect the presence of unit roots or stationarity.

The table contains seven columns, each representing different variables and the corresponding test statistics for the two main tests: ADF and DFGLS. Several important findings about the stationarity of the variables can be derived from this table.

In the level domain, the variables are tested for stationarity without differentiation. The null hypothesis for these tests assumes the presence of a unit root (implying non-stationarity), while the alternative hypothesis suggests stationarity.

- Transport [OS2] and Production and Demand of Goods and Services [CE2] show stationarity in all two tests: ADF and DFGLS.
- Private Consumption Indicator [PCI] shows stationarity in all three tests, except for the ADF test with drift and a deterministic trend (τ_{τ}).
- Volume and Values of Goods and Services [CE1] and Gross Domestic Product [GDP] are stationary as shown by the ADF test with drift and a deterministic trend (τ_{μ}).
- Personal and Household Goods [OS1] is found to be stationary by the ADF test with drift and a deterministic trend (τ_{τ}), and the DFGLS test with drift and a deterministic trend (τ_{τ}).

- Consumer Confidence Sentiment Index [CCSI] is identified as stationary by the DFGLS test with a drift term ($\tau\mu$).

Conversely, all other variables in the level range show non-stationarity, which means that the null hypotheses of the unit root tests are not refuted.

In the first differentiation, the variables are checked for stationarity after they have been differentiated once (with the Δ -operator). Differentiation is a common technique to convert non-stationary variables into stationary ones. The results of both tests show that Volume and Values of Goods and Services [CE1] is non-stationary. Also, Transport [OS2], Production and Demand of Goods and Services [CE2], Consumer Confidence Sentiment Index [CCSI], Gross Domestic Product [GDP] and Private Consumption Indicator [PCI] are classified as non-stationary. However, for all other variables, the p-value associated with the test statistic is smaller than the significance level. This means that the null hypothesis of a unit root is rejected and these variables are classified as stationary in their first-difference form.

Table 4.8: Unit Root and Stationarity Tests

	Test Statistics			
	$t\mu$	$t\tau$	$\tau\mu$	$\tau\tau$
Level				
OS1	-2.254 (0)	-4.265 (0)*	-1.206 (0)	-4.289 (0)*
OS2	-3.541 (0)*	-3.923 (0)*	-3.058 (0)*	-3.963 (0)*
CE1	-2.217 (1)	-2.209 (1)	-2.185 (1)*	-2.289 (1)
CE2	-7.731 (0)*	-7.689 (0)*	-7.752 (0)*	-7.769 (0)*
CCSI	-2.261 (3)	-2.263 (3)	-2.257 (3)*	-2.278 (3)
GDP	-2.422 (3)	-2.404 (3)	-2.436 (3)*	-2.439 (3)

PCI	-3.330 (0)*	-3.313 (0)	-3.346 (0)*	-3.348 (0)*
First Differences				
Δ OS1	-8.353 (1)*	-8.327 (1)*	-9.774 (0)*	-8.148 (1)*
Δ OS2	-10.505 (0)*	-10.447 (0)*	-8.492 (0)*	-9.743 (0)*
Δ CE1	-13.370 (0)*	-13.311 (0)*	-12.705 (0)*	-13.205 (0)*
Δ CE2	-8.655 (2)*	-8.600 (2)*	-8.464 (2)*	-8.649 (2)*
Δ CCSI	-8.535 (2)*	-8.483 (2)*	-8.588 (2)*	-8.592 (2)*
Δ GDP	-8.528 (2)*	-8.478 (2)*	-8.581 (2)*	-8.585 (2)*
Δ PCI	-8.110 (2)*	-8.059 (2)*	-8.161 (2)*	-8.161 (2)*

Notes: The t , τ and η statistics stand for ADF and DFGLS respectively. The subscript μ in the model allows for a drift term, while τ allows for a drift and a deterministic trend. See the main text for the labels. Asterisks (*) indicate that they are statistically significant at the 5 per cent level. The numbers in parentheses are the lag lengths. The asymptotic and finite sample critical values for ADF are from MacKinnon (1996). The DFGLS for the drift term (μ) follows the critical values of MacKinnon (1996), while the asymptotic distributions for the drift and deterministic trend (τ) are from Elliott et al. (1996, Table 1, p.825). Both the ADF and DFGLS tests examine the null hypothesis of a unit root versus the stationary alternative. Δ denotes the first difference operator.

In summary, the results of the ADF and DFGLS tests allow different interpretations. Overall, these tests indicate non-stationarity for many variables at their baseline level, while they signal stationarity for the same variables when examined in their first-difference form. Consequently, the ADF and DFGLS results suggest that the variables are characterised as stationary after differentiation and correspond to an integrated process of order one $I(1)$. This correspondence with a well-functioning stochastic process is confirmed by the significance levels and asterisks in Table 4.8.

4.4.2 Cointegration Test

After conducting unit root tests to determine the stationarity properties of individual time series variables, the next step is often to test whether these variables have a long-run relationship, a phenomenon known as cointegration. The cointegration test is a procedure that examines whether two or more non-stationary time series share a common long-run equilibrium relationship, signifying the parallel movements over time. This test is used to identify common trends and shocks between these time series and facilitates the estimation of the parameters of the cointegrating equation. Usually, the cointegration tests are performed following the unit root tests, as the results of the unit root tests determine the order of integration of the time series and thus influence the choice of the appropriate cointegration test and its critical values. In cases where the time series are integrated with integration order one $I(1)$, the Johansen tests for cointegration are often used.

Table 4.9 gives an overview of the results of the Johansen cointegration test applied to the online private consumption model. This test first evaluates different ranks indicating the number of cointegrating equations to determine the presence of cointegrating relationships within the data set. A cointegrating relationship means that a group of non-stationary time series shares a common stochastic trend, indicating a persistent equilibrium relationship between the variables. The test relies on two key statistics: the maximum eigenvalue ($\lambda_{\max}$) and the trace statistic. These statistics help evaluate the null hypothesis, which assumes r cointegrating vectors, against the alternative of $r+1$ cointegrating vectors. The test statistics are then compared to critical values at the 95% confidence level to determine their significance.

Table 4.9: Cointegration Test – Online Private Consumption Model

Null	Alternative	k=2 r=2			
		$\lambda_{\max}$		Trace	
		Unadjusted	95 per cent C.V.	Unadjusted	95 per cent C.V.
r = 0	r = 1	33.210	33.877	86.943*	69.819
r ≤ 1	r = 2	31.093*	27.584	53.733*	47.856
r ≤ 2	r = 3	13.904	21.132	22.640	29.797
r ≤ 3	r = 4	5.8639	14.265	8.7365	15.495
r ≤ 4	r = 5	2.8726	3.8415	2.8726	3.8415

Notes: Statistically significant values at the 5 per cent level are indicated by asterisks (*). “k” stands for lag length and “r” for the number of cointegrating vector(s). The critical values for the trace and maximum eigenvalue tests are from Johansen and Juselius (1990) and can be found in the tables of Osterwald-Lenum (1992).

These hypotheses are the null hypothesis (Ho) and the alternative hypothesis (Ha), which underlie the cointegration test. In this context, the null hypothesis states that the number of cointegrating vectors is less than or equal to the specified value of r, while the alternative hypothesis asserts that the number of cointegrating vectors is exactly equal to r. The statistic labelled $\lambda_{\max}$ (Unadjusted) denotes the test statistic for the maximum eigenvalue test ($\lambda_{\max}$) without any adjustments, while the statistic labelled Trace (Unadjusted) is the test statistic for the trace test without any modifications. The column "95 per cent C.V." contains critical values at the 5 per cent significance level that apply to both the maximum eigenvalue test and the trace test.

The trace test evaluates the sum of the eigenvalues against the critical values to determine the rank. For the hypothesis of no cointegration (rank 0), the test statistic is 86.943, which is above the critical value at the 0.05 significance level (69.819). Consequently, this

hypothesis is rejected. The situation is similar with the hypothesis of at most 1 cointegration (rank 1): The test statistic reaches 53.733, exceeding the critical value at the significance level of 0.05 (47.856), leading to the rejection of this hypothesis. Consequently, the trace test indicates the presence of 2 cointegrating equations at the 0.05 significance level.

The maximum eigenvalue test compares the largest eigenvalue with the critical values. For the hypothesis that there is no cointegration (rank 0), the test statistic of 33.210 falls below the critical value at the 0.05 significance level (33.877). Consequently, this hypothesis is not rejected. For the hypothesis of at most 1 cointegration (rank 1), on the other hand, the test statistic is 31.093 and thus above the critical value at the significance level of 0.05 (27.584). Consequently, this hypothesis is rejected. Essentially, the test for maximum eigenvalues does not provide support for cointegration at the 0.05 significance level.

Table 4.9 provides valuable insights into the cointegration of the online private consumption model. The null hypothesis of $r = 0$ is rejected by the trace statistic, but not by the $\lambda_{\max}$ statistic. This suggests the presence of at least one cointegrating vector within the model, although it is not conclusive evidence for the presence of more than one. Furthermore, the null hypothesis of $r \leq 1$ is rejected by both the $\lambda_{\max}$ and trace statistics. This means that there are at least two cointegrating vectors in the model, and we can safely claim that there are exactly two, since the null hypothesis of $r \leq 2$ is not challenged by either statistic.

In summary, the results indicate the presence of two potential cointegrating relationships between the variables within the model. While the trace test strongly suggests the presence of two cointegrating equations, the maximum eigenvalue test doesn't provide

sufficient evidence for cointegration. It is worth noting that in simulations with small samples, the trace test has shown better performance compared to the maximum eigenvalue test (Lütkepohl et al., 2001). When interpreting these results, it is important to consider the theoretical context of the variables and the significance of their coefficients. These tests provide valuable insights into potential long-run equilibrium relationships and short-run dynamics between the variables in this model and suggest the presence of a durable equilibrium relationship in the online private consumption model.

Since cointegration tests play an important role in the analysis of durable relationships between non-stationary time series variables, this section also conducts cointegration tests for a conventional private consumption model. As shown in Table 4.10, the Johansen test was performed using two lags and considering five possible cointegrating vectors. This test relies on both the maximum eigenvalue and trace statistics to evaluate the null hypothesis, which assumes r cointegrating vectors, as opposed to the alternative hypothesis, which suggests $r+1$ cointegrating vectors.

Table 4.10: Cointegration Test – Conventional Private Consumption Model

Null	Alternative	k=2 r=1			
		$\lambda_{\max}$		Trace	
		Unadjusted	95 per cent C.V.	Unadjusted	95 per cent C.V.
r = 0	r = 1	47.012*	33.877	87.540*	69.819
r ≤ 1	r = 2	18.475	27.584	40.529	47.856
r ≤ 2	r = 3	10.549	21.132	22.054	29.797
r ≤ 3	r = 4	7.9291	14.265	11.505	15.495
r ≤ 4	r = 5	3.5756	3.8415	3.5756	3.8415

Notes: Statistically significant values at the 5 per cent level are indicated by asterisks (*). “k” stands for lag length and “r” for the number of cointegrating vector(s). The

critical values for the trace and maximum eigenvalue tests are from Johansen and Juselius (1990) and can be found in the tables of Osterwald-Lenum (1992).

The trace test shows the trace statistics for various hypothetical ranks, a method of evaluating rank by comparing eigenvalues with critical values. In this case, the test indicates the presence of 1 cointegrating equation at the 0.05 significance level. Similar to the trace test, the maximum eigenvalue test determines the rank based on the largest eigenvalues and also indicates the presence of a cointegrating equation at the 0.05 significance level.

Given these results, it is evident that the null hypothesis of $r = 0$ can be safely rejected at the 5% significance level. Both the unadjusted lambda-max and trace statistics exceed their respective critical values, indicating the presence of at least one cointegrating vector between the variables of the model. However, the null hypothesis of $r \leq 1$ cannot be refuted as both statistics fall below their critical values. Consequently, there is no clear evidence for the presence of more than one cointegrating vector in the conventional private consumption model.

The chosen number of cointegrating vectors is therefore $r = 1$, a conclusion consistent with the results of both tests. This underlines the existence of a well-defined long-run equilibrium relationship between the variables within the conventional private consumption model.

4.4.3 VECM and Causality Test

After establishing cointegration and confirming a long-run relationship between variables, subsequent steps often include Granger Causality Tests and Error Correction Term (ECT). These tests provide further insight into the dynamics of the cointegrated system and provide information on how the variables influence each other over time.

In cases where two or more time series exhibit cointegration, it is imperative that Granger causality exists in at least one direction. Granger Causality Tests are useful to clarify both the direction and the strength of the causal relationship between the cointegrated time series. These tests examine whether the historical values of one variable provide valuable insights for predicting the behaviour of another variable. It is important to recognise that Granger causality works both ways: Variable A is considered a "cause" of variable B if past values of A contribute to the predictability of future values of B and vice versa.

When two or more time series exhibit cointegration, their deviations from long-term equilibrium become stationary and are suitable for use in an Error Correction Model (ECM). Typically, an ECM contains lagged differences of the cointegrated variables together with an Error Correction Term. This Error Correction Term serves as a measure of the speed of the adjustment process back to long-term equilibrium. If the deviations from equilibrium in the short run are significant, the significance of the error correction term underlines the existence of a mechanism that brings the system back to equilibrium.

Error! Not a valid bookmark self-reference. shows the results of the Granger Causality Test for the model of private online consumption, which contains five variables: Personal and Household Goods [OS1], Transport [OS2], Consumer Confidence Sentiment

Index [CCSI], Gross Domestic Product [GDP], and Private Consumption Indicator [PCI]. The Granger Causality Test is a statistical approach to determine whether a time series variable can improve the prediction of another time series variable based on its past values. This test is performed by estimating a two-lag Vector Error Correction Model (VECM) that includes a cointegration vector. The VECM comprehensively captures both short-run and long-run dynamics between variables. The short-term dynamics are represented by lagged differences in the variables, while the long-term dynamics are captured by the ECT, which quantifies the deviations from the equilibrium relationship.

Table 4.11: Granger Causality Results – Online Private Consumption Model

Dependent Variable	χ^2 -statistics					Coefficient of ECT
	Δ PCI	Δ OS1	Δ OS2	Δ CCSI	Δ GDP	
Δ PCI		1.018 (0.601)	2.119 (0.347)	6.786 (0.033)*	1.471 (0.479)	-0.708 (-4.841)*
Δ OS1	1.543 (0.462)		10.871 (0.004)*	0.558 (0.757)	0.255 (0.881)	-0.005 (-0.280)
Δ OS2	4.431 (0.109)	5.651 (0.059)		1.863 (0.394)	2.112 (0.348)	-0.023 (-0.790)
Δ CCSI	1.006 (0.605)	0.095 (0.954)	2.227 (0.328)		2.047 (0.359)	2.021 (3.567)
Δ GDP	0.015 (0.992)	0.181 (0.914)	1.119 (0.571)	6.068 (0.048)*		-0.630 (-4.625)*

Notes: The χ^2 -statistic tests the joint significance of the lagged values of the independent variables and the significance of the error correction term(s). Δ is the first different operator. Asterisks (*) indicate that they are statistically significant at the 5 per cent level. The numbers in brackets are the probability value.

Error! Not a valid bookmark self-reference. shows the results of the Granger Causality Test for the model of private online consumption, which contains five variables: Personal and Household Goods [OS1], Transport [OS2], Consumer Confidence Sentiment Index [CCSI], Gross Domestic Product [GDP], and Private Consumption Indicator [PCI]. The Granger Causality Test is a statistical approach to determine whether a time series variable can improve the prediction of another time series variable based on its past values. This test is performed by estimating a two-lag Vector Error Correction Model (VECM) that includes a cointegration vector. The VECM comprehensively captures both short-run and long-run dynamics between variables. The short-term dynamics are represented by lagged differences in the variables, while the long-term dynamics are captured by the ECT, which quantifies the deviations from the equilibrium relationship.

Table 4.11 illustrates the results of the Causality Test under the VECM. The χ^2 -statistics show significant results; Transport [OS2] is the causes of Personal and Household Goods [OS1]; Consumer Confidence Sentiment Index [CCSI] is the causes of both Private Consumption Indicator [PCI], and Gross Domestic Product [GDP] in the short run.

Examination of the ECT within the model shows that Personal and Household Goods [OS1] and Transport [OS2] play their important role in the short-run adjustment process to achieve long-run equilibrium in the online private consumption model. Specifically:

- The ECT coefficient for Private Consumption Indicator [PCI] shows a speed of adjustment of 0.708, which means that Private Consumption Indicator [PCI] takes about 2 months to fully adjust to the long-run equilibrium when the short-run adjustments are taken into account.

- In contrast, the ECT coefficient for Gross Domestic Product [GDP] shows a faster adjustment, accounting for 63% of the adjustment within a single month. Consequently, it takes about 2 months for Gross Domestic Product [GDP] to approach long-term equilibrium.

In short, Granger Causality Tests and Error Correction Tests provide valuable insights into the dynamics between variables within a cointegrated system. In the context of the online private consumption model, several notable Granger causality relationships emerge. As shown in Figure 4.1, Transport [OS2] exhibits Granger causality with Personal and Household Goods [OS1], while Consumer Confidence Sentiment Index [CCSI] exhibits Granger causality with both Gross Domestic Product [GDP] and Private Consumption Indicator [PCI]. Conversely, there is no substantial evidence of Granger causality from Personal and Household Goods [OS1], Gross Domestic Product [GDP] or Private Consumption Indicator [PCI] to other variables.

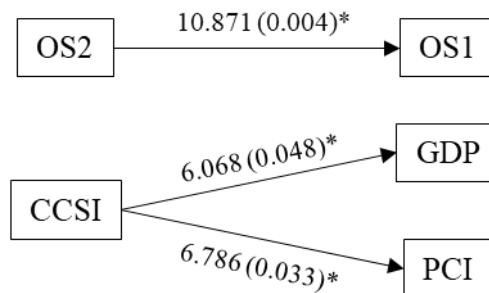


Figure 4.1: Short Run Causal Linkages – Online Private Consumption Model

After examining the cointegration dynamics of the system, it is advisable to examine the causal relationships between the variables. In particular, Production and Demand of

Goods and Services [CE2] turns out to be an endogenous variable, as indicated by the statistically significant ECT in the conventional private consumption model (as shown in **Error! Not a valid bookmark self-reference.**). This means that Production and Demand of Goods and Services [CE2] bears the main responsibility for short-term adjustments aimed at achieving long-term equilibrium.

Table 4.12 shows the results of the Granger Causality Test for the conventional private consumption model. This test can be used to determine whether past values of one time series have predictive power for future values of another time series, indicating possible causal relationships between variables. The test relies on the chi-square statistic, which evaluates the null hypothesis that the lagged values of the independent variables taken together have no significance for the dependent variable. In addition, the test includes the ECT, which measures the rate of adjustment towards long-run equilibrium within a cointegrated system.

After examining the cointegration dynamics of the system, it is advisable to examine the causal relationships between the variables. In particular, Production and Demand of Goods and Services [CE2] turns out to be an endogenous variable, as indicated by the statistically significant ECT in the conventional private consumption model (as shown in **Error! Not a valid bookmark self-reference.**). This means that Production and Demand of Goods and Services [CE2] bears the main responsibility for short-term adjustments aimed at achieving long-term equilibrium.

Table 4.12: Granger Causality Results – Conventional Private Consumption Model

χ^2 -statistics
77

Dependent Variable	Δ PCI	Δ CE1	Δ CE2	Δ CCSI	Δ GDP	Coefficient of ECT
Δ PCI		20.097 (0.000)*	48.646 (0.000)*	6.691 (0.035)*	6.360 (0.042)*	-0.783 (-5.535)
Δ CE1	0.377 (0.828)		2.522 (0.283)	9.612 (0.008)*	1.867 (0.393)	0.025 (1.112)*
Δ CE2	3.715 (0.156)	7.704 (0.021)*		1.234 (0.540)	6.265 (0.044)*	-0.204 (-4.786)*
Δ CCSI	0.055 (0.973)	5.489 (0.064)	6.600 (0.037)*		1.253 (0.535)	0.436 (0.748)
Δ GDP	4.142 (0.126)	26.058 (0.000)*	79.700 (0.000)*	3.946 (0.139)		-0.620 (-5.363)

Notes: The χ^2 -statistic tests the joint significance of the lagged values of the independent variables and the significance of the error correction term(s). Δ is the first different operator. Asterisks (*) indicate that they are statistically significant at the 5 per cent level. The numbers in brackets are the probability value.

The t-statistics associated with the lagged residuals are also statistically significant and show a negative trend within the model, which is consistent with Johansen's earlier results. The adjustment rate for Production and Demand of Goods and Services [CE2] is calculated to be 0.204, suggesting that it takes about 5 months for Production and Demand of Goods and Services [CE2] to move towards long-term equilibrium in the wake of these short-term adjustments.

Furthermore, the short-run Granger causality analysis shows that causal relationships exist between the variables within the conventional private consumption model under study, at a 5% significance level. Specifically, Volume and Values of Goods and Services [CE1] is identified as a causal factor for Private Consumption Indicator [PCI], Production and Demand of Goods and Services [CE2] and Gross Domestic Product [GDP]. Similarly,

Production and Demand of Goods and Services [CE2] acts as a causal factor for Private Consumption Indicator [PCI], Consumer Confidence Sentiment Index [CCSI] and Gross Domestic Product [GDP], while Consumer Confidence Sentiment Index [CCSI] causally influences both Private Consumption Indicator [PCI] and Volume and Values of Goods and Services [CE1]. Moreover, Gross Domestic Product [GDP] is found to causally influence both Private Consumption Indicator [PCI] and Production and Demand of Goods and Services [CE2]. These short-term relationships illustrate the interdependencies between the variables in the short run, and a summary of these causal relationships is shown in Figure 4.2.

Figure 4.2 shows a network of directional Granger causality between all pairs of variables within this model, except for the unidirectional causal relationship mentioned earlier. There is also a bidirectional causal relationship between Gross Domestic Product [GDP] and Production and Demand of Goods and Services [CE2] ($GDP \leftrightarrow CE2$).

A closer look at the causal factors within the system reveals an examination of causal chains. This means following a path from Volume and Values of Goods and Services [CE1] to Gross Domestic Product [GDP] and then to Production and Demand of Goods and Services [CE2] ($CE1 \rightarrow GDP \rightarrow CE2$) or vice versa from Volume and Values of Goods and Services [CE1] to Production and Demand of Goods and Services [CE2] and then to Gross Domestic Product [GDP] ($CE1 \rightarrow CE2 \rightarrow GDP$). An indirect relationship arises from Volume and Values of Goods and Services [CE1] to Gross Domestic Product [GDP] to Private Consumption Indicator [PCI] ($CE1 \rightarrow GDP \rightarrow PCI$). Furthermore, Volume and Values of Goods and Services [CE1] exerts a causal influence on Private Consumption Indicator [PCI],

with this causal effect cascading across the intervening variables (CE1-> CE2-> CCSI-> PCI).

In addition, there is another indirect relationship as Production and Demand of Goods and Services [CE2] exerts an influence on Private Consumption Indicator [PCI] via the intervening variable Consumer Confidence Sentiment Index [CCSI] (CE2-> CCSI-> PCI). Furthermore, Volume and Values of Goods and Services [CE1], Production and Demand of Goods and Services [CE2] and Consumer Confidence Sentiment Index [CCSI] are interconnected within the system in a circular causality pattern.

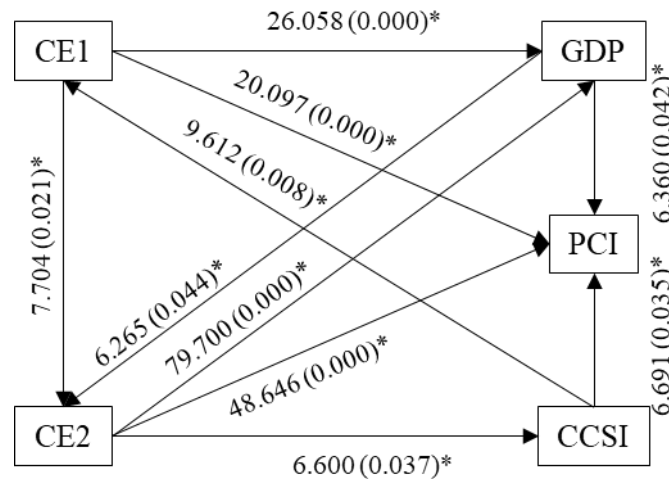


**Figure 4.2: Short Run Causality Direction –
Conventional Private Consumption Model**

4.4.4 VAR Variance Decomposition

After estimating a VECM and understanding the short-run dynamics and long-run equilibrium relationships within a cointegrated system, conducting a variance decomposition

analysis is a valuable way to further investigate the contribution of each variable to the variance of the prediction errors over time. Variance decomposition quantifies how much of the variance in the forecast error of one variable can be attributed to shocks or innovations in other variables. This analysis is conducted over different forecast horizons and provides a dynamic perspective on how the contributions of variables evolve over time. This approach is invaluable in understanding both short-term and long-term relationships and in shedding light on the evolution of the influence of variables over different time periods. Furthermore, this analysis supports decision-making, policy formulation, and a deeper understanding of the underlying system dynamics.

Table 4.13 shows the variance decomposition results for prediction horizons of 1, 4, 8, 12, 20, 30, 40 and 50 months. The table shows the predicted variances for five key variables, namely Personal and Household Goods [OS1], Transport [OS2], the consumer confidence sentiment index [CCSI], the macroeconomic variables comprising gross domestic product [GDP], and private consumption expenditure indicator [PCI]. The entries in the table represent percentages illustrating the proportion of deviations in the return forecasts that are due to random innovations.

Table 4.13 provides a detailed breakdown of the variances of the prediction errors over a 50-month horizon for each variable. These results provide insight into the dynamics within the model of private online consumption:

Table 4.13: Variance Decomposition – Online Private Consumption Model

Percentage of variations in Horizon (Month)	Due to Innovation in:					
	Δ PCI	Δ OS1	Δ OS2	Δ CCSI	Δ GDP	Δ CU
Months Relative Variance in: Δ PCI						
1	100.000	0.000	0.000	0.000	0.000	0.000
4	46.894	18.718	22.520	9.382	2.487	53.107
8	29.824	13.806	25.617	23.008	7.746	70.177
12	28.040	13.001	26.673	23.915	8.371	71.960
20	25.880	12.403	27.617	25.130	8.969	74.119
30	24.830	12.088	28.083	25.726	9.273	75.170
40	24.301	11.929	28.318	26.027	9.426	75.700
50	23.984	11.833	28.458	26.207	9.518	76.016
Months Relative Variance in: Δ OS1						
1	14.632	85.368	0.000	0.000	0.000	14.632
4	9.093	74.380	13.894	2.427	0.206	25.620
8	11.120	73.614	12.942	2.054	0.270	26.386
12	11.333	74.470	12.435	1.538	0.224	25.530
20	11.576	75.049	12.112	1.104	0.158	24.950
30	11.720	75.387	11.917	0.854	0.122	24.613
40	11.796	75.565	11.814	0.722	0.103	24.435
50	11.844	75.676	11.749	0.639	0.091	24.323
Months Relative Variance in: Δ OS2						
1	18.517	50.646	30.838	0.000	0.000	69.163
4	8.743	34.200	51.321	5.220	0.516	48.679
8	9.625	30.287	54.588	5.002	0.498	45.412
12	9.585	29.612	56.146	4.292	0.366	43.855
20	9.528	28.723	57.711	3.803	0.235	42.289
30	9.509	28.247	58.570	3.511	0.163	41.430
40	9.499	27.995	59.024	3.357	0.124	40.975
50	9.493	27.840	59.304	3.262	0.101	40.696
Months Relative Variance in: Δ CCSI						
1	3.746	1.131	0.017	95.106	0.000	4.894
4	3.722	9.622	1.107	85.174	0.375	14.826

8	6.416	13.076	3.713	72.951	3.844	27.049
12	6.724	13.935	4.420	69.983	4.938	30.017
20	7.194	14.979	5.141	66.766	5.919	33.233
30	7.453	15.559	5.555	64.941	6.492	35.059
40	7.593	15.869	5.776	63.962	6.800	36.038
50	7.679	16.062	5.914	63.353	6.991	36.646

Months Relative Variance in: Δ GDP

1	71.203	0.204	0.041	5.782	22.771	77.230
4	29.168	15.687	18.164	4.045	32.935	67.064
8	17.170	11.014	20.435	7.929	43.452	56.548
12	15.626	10.178	21.059	7.147	45.991	54.010
20	13.938	9.552	21.670	6.798	48.042	51.958
30	13.119	9.227	21.967	6.611	49.076	50.924
40	12.708	9.064	22.115	6.518	49.594	50.405
50	12.462	8.966	22.205	6.463	49.905	50.096

Notes: The numbers in the first column indicate the number of months. All other values are estimated and rounded to three decimal places - there may be rounding errors which may slightly affect the accuracy of the percentage decomposition in certain cases. The last column gives the percentage of the variance of the prediction errors for each variable that is explained in total by the other variables. The numbers in bold denote the effects of their own shocks.

The results show that Personal and Household Goods [OS1 as an exogenous variable appears to be the most exogenous (independent) variable in the system. Over a period of 50 months, a significant part (76%) of its predictive variance remains unexplained by the other variables in the model.

In contrast, Private Consumption Indicator [PCI] turns out to be the most interactive and endogenous (dependent) variable within the system. The cumulative shocks to which Private Consumption Indicator [PCI] is exposed over a period of 50 months are primarily

influenced by external factors, which account for about 76%. Specifically, these external shocks are composed of contributions from Personal and Household Goods [OS1] - 12%, Transport [OS2] - 28%, Consumer Confidence Sentiment Index [CCSI] - 26% and Gross Domestic Product [GDP] - 10%.

The results suggest a robust pattern of direct causality, with Transport [OS2] and Consumer Confidence Sentiment Index [CCSI] exerting a significant influence on Private Consumption Indicator [PCI]. This pattern is consistent with the observed cumulative shocks and their origins, and highlights the importance of Transport [OS2] and Consumer Confidence Sentiment Index [CCSI] for the fluctuations in Private Consumption Indicator [PCI].

In summary, the variance decomposition results in Table 4.13 underline the different degree of exogeneity and endogeneity of the model variables. Personal and Household Goods [OS1] is characterised by being less influenced by internal factors, while Private Consumption Indicator [PCI] is significantly influenced by external shocks, especially Transport [OS2] and Consumer Confidence Sentiment Index [CCSI]. These results contribute to a deeper understanding of the dynamics and causal relationships of the system.

Table 4.14 presents a comprehensive variance decomposition analysis covering forecast horizons from 1 to 50 months for the variables in the conventional private consumption model. The variables Volume and Values of Goods and Services [CE1], CE2, Consumer Confidence Sentiment Index [CCSI], Gross Domestic Product [GDP], and Private Consumption Indicator [PCI] are considered, with the following observations and results being noteworthy:

Table 4.14: Variance Decomposition – Conventional Private Consumption Model

Percentage of variations in Horizon (Month)	Due to Innovation in:					
	Δ PCI	Δ CE1	Δ CE2	Δ CCSI	Δ GDP	Δ CU
Months Relative Variance in: Δ PCI						
1	100.000	0.000	0.000	0.000	0.000	0.000
4	69.836	9.924	11.883	2.010	6.347	30.164
8	45.521	8.324	26.131	2.656	17.367	54.478
12	38.347	8.500	29.551	2.514	21.089	61.654
20	30.800	8.514	33.279	2.434	24.972	69.199
30	26.459	8.529	35.427	2.386	27.199	73.541
40	24.117	8.537	36.586	2.359	28.400	75.882
50	22.653	8.542	37.311	2.343	29.151	77.347
Months Relative Variance in: Δ CE1						
1	8.825	91.175	0.000	0.000	0.000	8.825
4	17.144	73.089	4.115	5.153	0.499	26.911
8	13.250	73.427	3.365	8.487	1.470	26.572
12	12.094	74.271	2.445	9.451	1.739	25.729
20	11.039	75.062	1.608	10.300	1.991	24.938
30	10.463	75.496	1.152	10.762	2.128	24.505
40	10.161	75.723	0.913	11.004	2.199	24.277
50	9.975	75.862	0.767	11.153	2.243	24.138
Months Relative Variance in: Δ CE2						
1	12.286	4.254	83.46	0.000	0.000	16.540
4	16.067	5.782	70.263	3.518	4.370	29.737
8	18.043	4.959	62.607	6.346	8.044	37.392
12	19.183	4.316	58.265	7.874	10.361	41.734
20	20.506	3.511	52.995	9.749	13.239	47.005
30	21.459	2.936	49.203	11.099	15.303	50.797
40	22.047	2.582	46.866	11.931	16.574	53.134
50	22.446	2.342	45.282	12.495	17.436	54.719
Months Relative Variance in: Δ CCSI						

1	5.870	0.747	33.979	59.404	0.000	40.596
4	4.008	1.389	40.967	53.549	0.087	46.451
8	3.170	1.203	42.019	53.552	0.056	46.448
12	2.913	1.177	42.366	53.501	0.043	46.499
20	2.671	1.140	42.702	53.457	0.030	46.543
30	2.545	1.120	42.874	53.437	0.024	46.563
40	2.480	1.110	42.963	53.427	0.020	46.573
50	2.440	1.104	43.017	53.420	0.018	46.579

Months Relative Variance in: Δ GDP

1	72.867	0.009	12.613	0.011	14.500	85.500
4	39.186	13.510	17.955	1.719	27.630	72.370
8	22.591	11.354	11.659	6.283	48.112	51.887
12	16.827	11.672	8.543	7.258	55.700	44.300
20	11.264	11.764	5.609	8.272	63.091	36.909
30	8.165	11.824	3.973	8.834	67.204	32.796
40	6.524	11.856	3.107	9.132	69.381	30.619
50	5.509	11.875	2.571	9.316	70.729	29.271

Notes: The numbers in the first column indicate the number of months. All other values are estimated and rounded to three decimal places - there may be rounding errors which may slightly affect the accuracy of the percentage decomposition in certain cases. The last column gives the percentage of the variance of the prediction errors for each variable that is explained in total by the other variables. The numbers in bold denote the effects of their own shocks.

Volume and Values of Goods and Services [CE1] turns out to be the most exogenous variable within the system for the conventional model of private consumption. Over a 50-month period, a remarkable 76% of its forecast variance remains unexplained by the other variables in the model.

Conversely, Private Consumption Indicator [PCI] stands out as the most interactive and endogenous variable within the system. The cumulative shocks faced by Private Consumption Indicator [PCI] over a 50-month period are primarily due to external influences, which account for about 77%. Specifically, these external shocks come from Volume and Values of Goods and Services [CE1] - 9%, Production and Demand of Goods and Services [CE2] - 37%, Consumer Confidence Sentiment Index [CCSI] - 2%, and Gross Domestic Product [GDP] - 29%, highlighting the significant role of Production and Demand of Goods and Services [CE2], and Gross Domestic Product [GDP] in shaping Private Consumption Indicator [PCI].

The results show a robust pattern of direct causality, with Production and Demand of Goods and Services [CE2] and Gross Domestic Product [GDP] exerting a significant influence on Private Consumption Indicator [PCI]. This pattern is consistent with cumulative shocks and their origin and highlights the central role of Production and Demand of Goods and Services [CE2] and Gross Domestic Product [GDP] in driving fluctuations in Private Consumption Indicator [PCI].

The analysis shows a remarkable shift in the dynamics of the variables over time. In the short run (1-4 months), the movements of the variables are primarily determined by their own shocks. In the long run, however, the variables are increasingly interrelated, with interdependencies contributing to their fluctuations.

In summary, the results in Table 4.14 show the varying degrees of exogeneity and endogeneity of the variables in the model. Volume and Values of Goods and Services [CE1] exhibits greater independence, while Private Consumption Indicator [PCI] is significantly

affected by external shocks, especially those emanating from Production and Demand of Goods and Services [CE2] and Gross Domestic Product [GDP]. These results improve our understanding of the dynamics and causal relationships of the system and shed light on short- and long-term patterns.

4.5 Summary

In this chapter, the study focussed on the use of online search indicators in addition to conventional economic indicators to predict private consumption. Principal component analysis (PCA) was used to summarise and refine the information from different online search categories. The analysis involved an extraction of online search indicators from Google Trends data, with keywords categorised using the COICOP classification.

PCA was performed for each COICOP division and the factor scores loaded in the first component were calculated. The results were used to determine the variables for the online search indicator. The analysis revealed two main components representing online searches related to personal and household goods and transport. These components explained 83.06% of the total variance in the online search data.

Similarly, the conventional economic indicators were also subjected to PCA, with six variables selected for the analysis. The extracted components represented trade-related indicators and production-related indicators. These components explained 82.47% of the total variance of the conventional economic indicators.

The unit root tests (ADF and DFGLS) were used to assess stationarity. In terms of levels, several variables, including transport and production and demand for goods and services, showed stationarity in all three tests, while other variables showed mixed results. First-difference transformations were performed to achieve stationarity for most variables.

Cointegration tests, in particular the Johansen cointegration test, were then performed to determine long-run relationships between non-stationary variables. The model of private online consumption showed evidence of at least two cointegrating relationships, with the trace test being more meaningful than the maximum eigenvalue test. After cointegration, Granger Causality Tests were conducted to understand the short-term dynamics. In the online private consumption model, traffic was found to cause a causal Granger relationship with personal and household goods, while the consumer confidence index showed a causal Granger relationship with both the private consumption indicator and gross domestic product.

For the conventional model of private consumption, the Granger causality results indicated significant relationships between the variables, with production and demand for goods and services playing a key role in the short-term adjustments.

The Vector Error Correction Model (VECM) and the examination of the Error Correction Term (ECT) provided insights into the adjustment processes towards long-run equilibrium. The ECT coefficients for the different variables indicate the speed at which they adjust to reach equilibrium.

To summarise, the variance decomposition analysis carried out in this chapter provides valuable insights into the dynamics and causal relationships within the models of

online private consumption and conventional private consumption examined. For online private consumption, the variable personal and household goods [OS1] proves to be the most exogenous variable, with a considerable proportion (76%) of its predictive variance over a period of 50 months remaining unexplained by other variables. The private consumption indicator [PCI], on the other hand, proves to be the most interactive and endogenous variable, with the cumulative shocks coming primarily from external factors such as transport [OS2], the consumer confidence sentiment index [CCSI], and gross domestic product [GDP].

The conventional model of private consumption also shows different patterns of exogeneity and endogeneity. The volume and value of goods and services [CE1] stands out as the most exogenous variable, with 76% of its predictive variance over 50 months not explained by other variables. The private consumption indicator [PCI], on the other hand, remains highly endogenous, with the cumulative shocks being primarily determined by the production and demand of goods and services [CE2] and gross domestic product [GDP].

CHAPTER 5

CONCLUSIONS

5.1 Summary of Findings

In the age of the Internet, its indispensability for basic needs such as food, clothing, housing and transport is undeniable. Consequently, the amount of data collected on user behaviour is constantly growing. Recently, there has been a surge in research activity in this area. Google Trends, as a service offered by the widely used Google search system, has established itself as a benchmark for predictions in various areas such as private consumption, the housing market and the stock market. This popularity is due to its relative real-time capability. The analysis of online data provides an immediate understanding of market dynamics. For example, Choi and Varian (2012) studied automobile sales and unemployment benefits, Herzog (2014) on corporate asset and stock prices, while Toth and Hajdu (2012) on housing and automobile markets. Much of the research has focussed on predicting private consumption through Google Trends (Kholodilin et al., 2010; Vosen & Schmidt, 2011, 2012). This work highlights the accumulated evidence on the predictive value of search-based indicators.

Through PCA, this study extracts the online search indicators, the conventional economic indicators and applies the time series forecasting and analysis modules to investigate the relevant advantages of including online search information for forecasting private consumption. The construction of the online search indicator is based on the product subcategory on the Shopee website in Malaysia. For the construction of the search indicator,

the total amount of data is aggregated into 248 categories of time series information for product search, and then the composition is done by each COICOP division.

Table 5.1 shows the results of the online search indicators in Malaysia, showing that personal and household goods, which include items such as home appliances, footwear, clothing, health products, recreational, and sporting goods, consistently have higher search volume on PCA than categories such as transportation and food and beverages.

Table 5.1: Summary of Methods and Results for Online Search Indicators and Conventional Economic Indicators

Process	Method	Results
Online Search Indicators		
Step 1: Download data from Google Trends	Keyword search based on product subcategories on the Shopee website; region is Malaysia; time range is from 1/1/2016 to 31/3/2023, with corresponding Google category, in Google web search	248 data downloaded; sample size is 87 (monthly data from 1/1/2016 to 31/3/2023)
Step 2: Classified downloaded data	Based on COICOP 2018	9 COICOP divisions created, each division consisting of more than 2 data
Step 3: Extraction of the first component from each COICOP division	Using PCA for each COICOP division	9 components extracted (1 from each COICOP division)
Step 4: Extraction of online search indicators from 9 components corresponding to 9 COICOP division	Using PCA	2 Components (OS1, OS2) extracted

Conventional Economic Indicators		
Extraction of conventional economic indicators, using official data from the DOSM	Period from 1/1/2016 to 31/3/2023, monthly frequency, quarterly data must be converted to monthly, with PCA	2 Components (CE1, CE2) extracted

One possible explanation is that Malaysians show a greater propensity to search and purchase household goods online, including essential items such as home appliances, clothing and health products. These product categories are closely linked to their lifestyle and personal needs, so they are more likely to shop online due to convenience and the ability of the system in meeting individual preferences.

In addition, certain categories such as health, leisure, and sporting goods may see an increase in search volume at certain times of the year, such as at the start of a new fitness season or during the holiday season when people are actively searching for such products.

Based on the results for the conventional economic indicators, changes in the volume of imports and exports in trade reflect shifts in supply and demand and provide a comprehensive picture of economic conditions. The producer price index reflects input costs for production, while imports of consumer goods provide information on procurement strategies and possible bottlenecks in the supply chain. These indicators reflect the quantity and value of food and services exchanged between trading partners, which means that the market is competitive and the economy is more open.

In the part of the study dealing with the prediction of private consumption using time series analysis, a series of tests were conducted. First, the results of the ADF and DFGLS

unit root test showed that all seven variables (“Online search indicator 1 - Personal and Household Goods” [OS1], “Online search indicator 2 - Transport” [OS2], “Conventional economic indicator 1 - volume and value of goods and services exchanged between trading partners” [CE1], “Conventional economic indicator 2 - Production and demand of goods and services within the country” [CE2], “Consumer confidence sentiments index” [CCSI], “Gross domestic product” [GDP], and “Private final consumption indicator” [PCI]) are all non-stationary at levels (have unit roots); but become stationary after differentiation once, since the p-value associated with a test statistic is smaller than the significance level. Looking at the lag lengths used in the test, most variables seem to behave according to the stochastic process I (1) after one-time differentiation.

Second, the Johansen cointegration test shows the trace and maximum eigenvalues performed to determine the order of integration; in the cointegration test of the online private consumption model, the null hypothesis $r \leq 1$ is rejected by both the $\lambda_{\max}$ and trace statistics, but it is also clear that there are at least two cointegrating vectors in the model, as the trace statistics are larger than the corresponding critical values, which there are exactly two long-run relationships between the variable and the online private consumption model. In the cointegration test of the conventional private consumption model, the result shows that the null hypothesis of $r = 0$ is rejected as both the unadjusted lambda max and trace statistics are greater than the corresponding critical values, so the cointegrating vectors are $r = 1$, which is consistent with both tests. This means that there is a clear long-run equilibrium relationship between the variables in the conventional private consumption model.

Third, cointegration implies a long-term relationship between variables, but deviations from this relationship (errors) can occur in the short term. Error Correction Tests

and the analysis of an ECM provide information about the adjustment process towards the long-run equilibrium when short-run deviations occur. In the online private consumption model, two long-run relationships are found in the cointegration test, but the results of the ECT coefficients and their t-ratio are negative and significant, with only two satisfying the conditions, namely the variables Personal and Household Goods [OS1] and Transport [OS2], which means that these variables adjust to restore the long-run equilibrium when there is a deviation from it. In the conventional private consumption model, the ECT coefficients are positive and non-significant, and only one of the ECT variables satisfies the ECT condition.

Fourth, analyse the results of the Granger Causality Test to understand the direction of both the short-term and long-term causal relationships between the variables. There are some significant Granger causality relationships between the variables in the online private consumption model; in the short run, Transport [OS2] Granger causes Personal and Household Goods [OS1]; and Consumer Confidence Sentiment Index [CCSI] Granger causes Gross Domestic Product [GDP] and Private Consumption Indicator [PCI]. For the conventional private consumption model, the results of the Granger Causality Test for the variables show bidirectional causality leading from Production and Demand of Goods and Services [CE2] to Gross Domestic Product [GDP] and from Gross Domestic Product [GDP] to Production and Demand of Goods and Services [CE2]. For the indirect relationship, the results of the Granger Causality Test indicate that Production and Demand of Goods and Services [CE2], Consumer Confidence Sentiment Index [CCSI] and Gross Domestic Product [GDP] can act as mediators in the relationship. The result shows that Volume and Values of Goods and Services [CE1], Production and Demand of Goods and Services [CE2], Consumer Confidence Sentiment Index [CCSI] and Gross Domestic Product [GDP] are good

exogenous variables and Private Consumption Indicator [PCI] is a good endogenous variable. However, this does not mean that there is a true causal relationship between these variables as there may be other factors that influence them.

Finally, a variance decomposition analysis was conducted to shed light on the respective contributions of each variable to the variances of the other variables' prediction errors. We examined these results across different forecast horizons so that we could identify both short-term and long-term dynamics within the system. In the online private consumption model, our results showed that Personal and Household Goods [OS1] proved to be the most exogenous variable, accounting for a remarkable 76% of the variance through its own shocks. Conversely, Private Consumption Indicator [PCI] proved to be the most interactive and endogenous variable within the system, explaining 24% of the variance through its own shocks, which was particularly evident after the 50-month horizon. Comparing these results with the conventional private consumption model, a similar pattern emerges. Volume and Values of Goods and Services [CE1] was identified as the most exogenous variable, still explaining 76% of the variance through its own shocks. Once again, Private Consumption Indicator [PCI] showed its importance as the most interactive and endogenous variable, explaining 24% of the variance through its own shocks, which is particularly remarkable after the 50-month horizon. In essence, these comparative results suggest that both models have comparable predictive power, which is particularly evident after the 50-month mark. In other words, given the variance decompositions, the online private consumption model could be a viable alternative or complement to the conventional private consumption model.

5.2 Implications of the Study

The study of private consumption forecasts has several implications that can benefit different stakeholders:

Understanding private consumption enables companies and retailers to optimise their operations by aligning production, inventories and marketing activities with expected consumer demand, taking a long-term perspective on Private Consumption Indicator [PCI], Gross Domestic Product [GDP] and the Production and Demand of Goods and Services [CE2]. This helps them to minimise over or understocking of goods, reduce costs and improve overall profitability, while the Consumer Confidence Sentiment Index [CCSI] is an exogenous variable for Gross Domestic Product [GDP], the Volume and Value of Goods and Services [CE1] and the Private Consumption Indicator [PCI] in the short term. It also helps in avoiding overproduction and waste and promoting sustainability through resource consumption minimisation and the reduction of environmental impacts associated with unnecessary production and distribution. Analysing forecasts of private consumption helps supply chain managers to anticipate fluctuations in demand and adjust procurement, production and distribution processes accordingly. This leads to efficient resource allocation and fewer disruptions in the supply chain.

If the Consumer Confidence Sentiment Index [CCSI] exhibits Granger causality for the Personal Consumption Indicator [PCI], companies that use consumer forecasts effectively will have a competitive advantage because they can stay ahead of market trends, make informed strategic decisions and offer products and services that consumers demand when they need them. Retailers and marketers can use this information to tailor products,

services, and marketing strategies to consumer needs and trends, with the Personal and Household Goods Indicator [OS1] and the Volume and Value of Goods and Services Indicator [CE1] being the most exogenous after 50 months. Understanding future consumer demand through forecasting can help companies innovate and develop new products that meet changing consumer preferences, which can lead to increased market share and competitiveness.

Estimating private consumption provides insights into consumer behaviour that can serve as a basis for economic, fiscal, and monetary policy measures. Using the Consumer Confidence Sentiment Index [CCSI] and Gross Domestic Product [GDP] as indirect or mediating variables allows governments to adjust taxes, interest rates, and stimulus packages based on expected changes in private consumption to promote economic stability and growth. The study of private consumption forecasts has major implications for economic stability, business profitability, efficient resource allocation, informed decision-making, and increased consumer satisfaction.

5.3 Limitations of the Study

The study uses a time series methodology to analyse search indicators generated by keyword phrases. These indicators offer the possibility of information accumulation in real time. However, the study does not currently use the immediacy of information retrieval to effectively capture the relevance of search terms and explain the corresponding search behaviour. The correlation of search behaviour involves different patterns of information retrieval, which is a technical challenge that has not yet been addressed in this study.

Currently, the study focuses on the construction of consumption-based search indicators. This construction involves the use of product items sourced from e-commerce websites. However, this approach encounters the difficulty of comprehensively capturing both product and service categories. The use of national accounts categories to classify private consumption and create search indices poses a number of problems. In particular, certain consumption categories, such as restaurants and hotels, may not correspond exactly with consumer usage patterns. The large scale required for this classification is beyond the capabilities of data collection within the timeframe set for the study. It is expected that these challenges will be further explored and assessed in future research.

5.4 Recommendations and Directions for Future Research

In the present study, search terms derived from the product categories on the Shopee website were used. However, due to the time constraints of the study, it was not possible to evaluate the predictive performance with different keyword modifications. Future research efforts should consider the use of different keyword construction models to investigate the predictive power of search indicators that encompass different components of private consumption.

This study is currently limited by its study period, which prevents an in-depth treatment of strong exogeneity. Future research should incorporate interpretations of exogeneity to extend the analysis of the benefits of private consumption forecasting indicators derived from online data. These insights can be used to evaluate business

strategies or policies, facilitating the establishment of key benchmarks and recommendations.

In this study, PCA is used to extract indicators. For future research, it is recommended to use a wider range of analytical techniques, including ridge regression and LASSO (Least Absolute Shrinkage and Selection Operator), to investigate and compare different composite measures of predictive validity.

The Private Consumption Prediction Indicator (database) is currently integrated with the Online Search Indicator using monthly data for the target from 2016, resulting in a limited sample size. For future research, the use of a mixed-frequency measurement approach, such as Giannone et al., (2008) is recommended. Incorporating high-frequency weekly (or even daily) data to predict low-frequency seasonal data, like private consumption, can unlock the potential of online data to provide real-time information benefits. This approach facilitates the use of timely updates to effectively improve forecasting capabilities.

As a recommendation for MSME development, capacity building programmes are often implemented that focus on improving the skills and capabilities of MSMEs, including training programmes, workshops, and mentoring initiatives to improve their competitiveness. MSME access can be initiated in national and international markets, for example through trade missions, export incentives and assistance with regulatory compliance and certification procedures, as well as innovation and entrepreneurship initiatives programmes to promote a culture of innovation and entrepreneurship, including incubators, accelerators, and start-up support programmes.

Policy proposals and recommendations for the government include the implementation of various financial assistance programmes and grants to support MSMEs in areas such as technology adoption, encouraging MSMEs to adopt digital technologies and innovation. These are often held through incentives and support for the implementation of Industry 4.0 practices, with the main focus on digital transformation and innovation, as well as measures to encourage companies to adopt digital technologies and improve their competitiveness. There are efforts to simplify legislation to streamline regulatory processes and reduce bureaucratic hurdles in order to make it easier for MSMEs to operate and comply with regulations. In times of economic downturn or crisis, governments can launch stimulus packages to inject funds into the economy to support businesses and consumers

For industrial development, development of clusters supports the creation of industrial clusters to promote cooperation, networking, and resource sharing among MSMEs in specific sectors. Support for research and development initiatives also promotes innovation and improves competitiveness. Investing in infrastructure projects on the other hand can boost economic activity, create more job opportunities, and produce a positive impact on private consumption.

REFERENCES

- Ambad, S. N. A., Andrew, J. V., & Awang Amit, D. H. D. (2020). Growth challenges of SMEs: empirical evidence in Sabah, Malaysia. *ASEAN Entrepreneurship Journal*, 6(1), 8-14.
- Anttonen, J. (2018). Nowcasting the unemployment rate in the EU with seasonal BVAR and Google search data. *ETLA Working Papers*, No. 62.
- Artola, C., Pinto, F., & de Pedraza García, P. (2015). Can internet searches forecast tourism inflows? *International Journal of Manpower*, 36(1), 103-116. <https://doi.org/10.1108/IJM-12-2014-0259>
- Askatas, N., & Zimmermann, K. (2009). Google Econometrics and Unemployment Forecasting. *Applied Economics Quarterly (formerly: Konjunkturpolitik)*, 55(2), 107-120. <https://doi.org/10.3790/aeq.55.2.107>
- Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, 46, 454-464. <https://doi.org/10.1016/j.tourman.2014.07.014>
- Bartels, J., & Onwezen, M. C. (2014). Consumers' willingness to buy products with environmental and ethical claims: the roles of social representations and social identity. *International Journal of Consumer Studies*, 38(1), 82-89. <https://doi.org/10.1111/ijcs.12067>
- Boik, A., Greenstein, S., & Prince, J. (2017). *The empirical economics of online attention*. Kelley School of Business Research Paper No. 16-57. <https://doi.org/10.2139/ssrn.2807046>
- CEIC Data. (2023). *Malaysia Private Consumption: % of GDP*. Retrieved August 20, 2023, from <https://www.ceicdata.com/en/indicator/malaysia/private-consumption--of-nominal-gdp>
- Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, 88(s1), 2-9. <https://doi.org/10.1111/j.1475-4932.2012.00809.x>
- Comscore. (2023). *Top 50 Multi-Platform Properties (Desktop and Mobile)*. Retrieved August 20, 2023, from https://www.comscore.com/Insights/Rankings?utm_campaign=APAC_MY_SEP2022_MR_MMXX_MPAUG2022&utm_medium=email&utm_source=comscore_elq_APAC_MY_SEP2022_MR_MMXX_MPAUG2022&elqCampaignID=9730&cs_edgescape_cc=MY&elqTra

ckId=57ab20a0923242bca2fb054822ea3679&elq=10af49ba57964adba30c9615c20832be&elqaid=14671&elqat=1&elqCampaignId=9730#content_0/

- Costola, M., Iacopini, M., & Santagiustina, C. R. M. A. (2021). Google search volumes and the financial markets during the COVID-19 outbreak. *Finance Research Letters*, 42, 101884. <https://doi.org/10.1016/j.frl.2020.101884>
- Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *The Journal of Finance*, 66(5), 1461-1499. <https://doi.org/10.1111/j.1540-6261.2011.01679.x>
- DataReportal. (2023a). *Digital 2023 Global Overview Report*. <https://datareportal.com/reports/digital-2023-global-overview-report>
- DataReportal. (2023b). *Digital 2023 Malaysia*. <https://datareportal.com/reports/digital-2023-malaysia>
- Department of Statistics Malaysia. (2023, May 31). *ICT Use and Access by Individuals and Households Survey Report 2022*. <https://www.dosm.gov.my/portal-main/release-content/82c59a73-8b80-11ed-96a6-1866daa77ef9>
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74(366a), 427-431. <https://doi.org/10.1080/01621459.1979.10482531>
- Dritsakis, N. (2004). Defense spending and economic growth: an empirical investigation for Greece and Turkey. *Journal of Policy Modeling*, 26(2), 249-264. <https://doi.org/10.1016/j.jpolmod.2004.03.011>
- Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica*, 64(4), 813-836. <https://doi.org/10.2307/2171846>
- Enders, W. (2015). *Applied econometric time series* (4th ed.). John Wiley & Sons.
- European Central Bank. (2018). *Private consumption and its drivers in the current economic expansion*. Retrieved August 20, 2023 from https://www.ecb.europa.eu/pub/economic-bulletin/articles/2018/html/ecb.ebart201805_03.en.html
- Fondeur, Y., & Karamé, F. (2013). Can Google data help predict French youth unemployment? *Economic Modelling*, 30, 117-125. <https://doi.org/10.1016/j.econmod.2012.07.017>
- Fuellhart, K. G., & Glasmeier, A. K. (2003). Acquisition, assessment and use of business information by small- and medium-sized businesses: a demand perspective. *Entrepreneurship & Regional Development*, 15(3), 229-252. <https://doi.org/10.1080/0898562021000011197>

- Gausden, R., & Hasan, M. S. (2018). An assessment of the contribution of consumer confidence towards household spending decisions using UK data. *Applied Economics*, 50(12), 1395-1411. <https://doi.org/10.1080/00036846.2017.1363859>
- Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4), 665-676. <https://doi.org/10.1016/j.jmoneco.2008.05.010>
- Gonzalo, J. (1994). Five alternative methods of estimating long-run equilibrium relationships. *Journal of Econometrics*, 60(1), 203-233. [https://doi.org/10.1016/0304-4076\(94\)90044-2](https://doi.org/10.1016/0304-4076(94)90044-2)
- Götz, T. B., & Knetsch, T. A. (2019). Google data in bridge equation models for German GDP. *International Journal of Forecasting*, 35(1), 45-66. <https://doi.org/10.1016/j.ijforecast.2018.08.001>
- Granger, C. W. J. (1988). Some recent development in a concept of causality. *Journal of Econometrics*, 39(1), 199-211. [https://doi.org/10.1016/0304-4076\(88\)90045-0](https://doi.org/10.1016/0304-4076(88)90045-0)
- Hamid, A., & Heiden, M. (2015). Forecasting volatility with empirical similarity and Google Trends. *Journal of Economic Behavior & Organization*, 117, 62-81. <https://doi.org/10.1016/j.jebo.2015.06.005>
- Herzog, B. (2014). Google's search data and its application in finance. *International journal of economics, finance and management sciences*, 2(1), 1-7. <https://doi.org/10.11648/j.ijefm.20140201.11>
- Hutchinson, K., Fleck, E., & Lloyd-Reason, L. (2009). An investigation into the initial barriers to internationalization. *Journal of Small Business and Enterprise Development*, 16(4), 544-568. <https://doi.org/10.1108/14626000911000910>
- International Monetary Fund. (2022, October). World Economic Outlook: Countering the Cost-of-Living Crisis. <https://www.imf.org/en/Publications/WEO/Issues/2022/10/11/world-economic-outlook-october-2022>
- Jocumsen, G. (2004). How do small business managers make strategic marketing decisions? *European Journal of Marketing*, 38(5/6), 659-674. <https://doi.org/10.1108/03090560410529277>
- Johansen, S., & Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration—with appucations to the demand for money. *Oxford Bulletin of Economics and statistics*, 52(2), 169-210.

- Juhro, S. M., & Iyke, B. N. (2020). Consumer confidence and consumption expenditure in Indonesia. *Economic Modelling*, 89, 367-377. <https://doi.org/10.1016/j.econmod.2019.11.001>
- Jun, S.-P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, 130, 69-87. <https://doi.org/10.1016/j.techfore.2017.11.009>
- Kholodilin, K. A., Podstawski, M., & Siliverstovs, B. (2010). Do Google searches help in nowcasting private consumption? A real-time evidence for the US. *KOF Swiss Economic Institute Working Paper No. 256, DIW Berlin Discussion Paper No. 997*. <https://doi.org/10.2139/ssrn.1615453>
- Koen, V., Asada, H., Nixon, S., Rahuman, M. R. H., & Mohd Arif, A. Z. (2017). Malaysia's economic success story and challenges. *OECD Economics Department Working Papers*, No. 1369. <https://doi.org/10.1787/cf7fddf2-en>
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1), 159-178. [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
- Lee, S. H., & Hoffman, K. D. (2015). Learning the ShamWow: Creating Infomercials to Teach the AIDA Model. *Marketing Education Review*, 25(1), 9-14. <https://doi.org/10.1080/10528008.2015.999586>
- Lütkepohl, H., Saikkonen, P., & Trenkler, C. (2001). Maximum eigenvalue versus trace tests for the cointegrating rank of a VAR process. *The Econometrics Journal*, 4(2), 287-310. <https://doi.org/10.1111/1368-423x.00068>
- Lybaert, N. (1998). The Information Use in a SME: Its Importance and Some Elements of Influence. *Small Business Economics*, 10(2), 171-191. <https://doi.org/10.1023/A:1007967721235>
- MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics*, 11(6), 601-618. [https://doi.org/10.1002/\(SICI\)1099-1255\(199611\)11:6<601::AID-JAE417>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1099-1255(199611)11:6<601::AID-JAE417>3.0.CO;2-T)
- Masih, A. M. M., & Masih, R. (2001). Dynamic Modeling of Stock Market Interdependencies: An Empirical Investigation of Australia and the Asian NICs. *Review of Pacific Basin Financial Markets and Policies*, 4(2), 235-264. <https://doi.org/10.1142/s0219091501000401>

- Masih, A. M. M., Masih, R., & Hasan, M. S. (1997). New evidence from an alternative methodological approach to the defence spending-economic growth causality issue in the case of mainland China. *Journal of Economic Studies*, 24(3), 123-140. <https://doi.org/10.1108/01443589710167347>
- Matsumoto, A., Matsumura, K., & Shirak, i. (2013). *Potential of Search Data in Assessment of Current Economic Conditions*. Bank of Japan Research Papers 2013-04-18. <https://ideas.repec.org/p/boj/bojron/13-e-0418.html>
- Mavragani, A., & Tsagarakis, K. P. (2016). YES or NO: Predicting the 2015 GReferendum results using Google Trends. *Technological Forecasting and Social Change*, 109, 1-5. <https://doi.org/10.1016/j.techfore.2016.04.028>
- Moat, H. S., Curme, C., Stanley, H. E., & Preis, T. (2014). Anticipating Stock Market Movements with Google and Wikipedia. 47-59. https://doi.org/10.1007/978-94-017-8704-8_4
- Mohamad, A., Mustapa, A. N., & Razak, H. A. (2021). An Overview of Malaysian Small and Medium Enterprises: Contributions, Issues, and Challenges. In B. S. Sergi & A. R. Jaaffar (Eds.), *Modeling Economic Growth in Contemporary Malaysia* (pp. 31-42). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-80043-806-420211004>
- Mustapa, M., Ponnusamy, R. R., & Ho, M. K. (2019). Forecasting prices of fish and vegetable using web scraped price micro data. *International Journal of Recent Technology and Engineering*, 7(5S), 251-256.
- O'Donnell, A. (2011). Small firm marketing: synthesising and supporting received wisdom. *Journal of Small Business and Enterprise Development*, 18(4), 781-805. <https://doi.org/10.1108/14626001111179802>
- Osterwald-Lenum, M. (1992). A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics. *Oxford Bulletin of Economics and statistics*, 54(3), 461-472. <https://doi.org/10.1111/j.1468-0084.1992.tb00013.x>
- Pavliček, J. (2014). *Google Econometrics: Unemployment in Visegrad Countries*. Bachelor thesis. Charles University in Prague.
- Pedrycz, W., & Chen, S.-M. (2013). *Social networks: a framework of computational intelligence* (Vol. 526). Springer.

- Perlin, M. S., Caldeira, J. F., Santos, A. A. P., & Pontuschka, M. (2017). Can We Predict the Financial Markets Based on Google's Search Queries? *Journal of Forecasting*, 36(4), 454-467. <https://doi.org/10.1002/for.2446>
- Phillips, P. C. B. (1991). Optimal Inference in Cointegrated Systems. *Econometrica*, 59(2), 283-306. <https://doi.org/10.2307/2938258>
- Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports*, 3(1), 1684. <https://doi.org/10.1038/srep01684>
- Rahman, N. A., Yaacob, Z., & Radzi, R. M. (2016). The challenges among Malaysian SME: A theoretical perspective. *World Journal of Social Sciences*, 6(3), 124-132.
- Sallis, J. E., Gripsrud, G., Olsson, U. H., & Silkoset, R. (2021). Factor Analysis. In *Research Methods and Data Analysis for Business Decisions: A Primer Using SPSS* (pp. 223-243). Springer International Publishing. https://doi.org/10.1007/978-3-030-84421-9_12
- Semrush. (2023). *Most Visited Websites by Traffic in Malaysia for all categories, July 2023*. Retrieved August 20, 2023, from <https://www.semrush.com/website/top/malaysia/all/>
- Shear, F., Ashraf, B. N., & Sadaqat, M. (2021). Are Investors' Attention and Uncertainty Aversion the Risk Factors for Stock Markets? International Evidence from the COVID-19 Crisis. *Risks*, 9(1), 2. <https://doi.org/10.3390/risks9010002>
- Similarweb. (2023). *Top Websites Ranking for eCommerce & Shopping in Malaysia*. Retrieved April 20, 2023, from <https://www.similarweb.com/top-websites/malaysia/category/e-commerce-and-shopping/>
- Simon, H. A. (1971). Designing organizations for an information-rich world. In M. Greenberger (Ed.), *Computers, communications, and the public interest* (pp. 37-72). Johns Hopkins Press.
- Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1), 1-48. <https://doi.org/10.2307/1912017>
- Sims, C. A. (1982). Policy Analysis with Econometric Models. *Brookings Papers on Economic Activity*, 1982(1), 107-164. <https://doi.org/10.2307/2534318>
- SME Corp. (2022). MSME Insights 2021: MSMEs towards Sustainable Recovery. https://www.smecorp.gov.my/images/Publication/MSME_Insights/SME_Insight_2021_1_English_Full_Version.pdf

- StatCounter. (2023). *Desktop & Mobile Browser Market Share in Malaysia - July 2023*. Retrieved August 20, 2023, from <https://gs.statcounter.com/browser-market-share/desktop-mobile/malaysia>
- SÜRÜCÜ, L., & MaslakÇI, A. (2020). Validity and reliability in quantitative research. *Business & Management Studies: An International Journal*, 8(3), 2694-2726. <https://doi.org/10.15295/bmij.v8i3.1540>
- Tehseen, S. (2014). Malaysian SMEs-An Overview of SMEs' Contributions and Challenges. 1st UniKL Business School Management and Entrepreneurship Conference (UBSMEC),
- Toth, I. J., & Hajdu, M. (2012). Google as a tool for nowcasting household consumption: estimations on Hungarian data. *Institute for Economic and Enterprise Research - HCCI*. <https://gvi.hu/research-details/47/google-as-a-tool-for-nowcasting-household-consumption:-estimations-on-hungarian-data>
- United Nations. (2018). Classification of Individual Consumption According to Purpose (COICOP) 2018. https://unstats.un.org/unsd/classifications/unsdclassifications/COICOP_2018_-_pre-edited_white_cover_version_-_2018-12-26.pdf
- Veldhuizen, S. v., Vogt, B., & Voogt, B. (2016). Internet searches and transactions on the Dutch housing market. *Applied Economics Letters*, 23(18), 1321-1324. <https://doi.org/10.1080/13504851.2016.1153785>
- Vicente, M. R., López-Menéndez, A. J., & Pérez, R. (2015). Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing? *Technological Forecasting and Social Change*, 92, 132-139. <https://doi.org/10.1016/j.techfore.2014.12.005>
- Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting*, 30(6), 565-578. <https://doi.org/10.1002/for.1213>
- Vosen, S., & Schmidt, T. (2012). A monthly consumption indicator for Germany based on Internet search query data. *Applied Economics Letters*, 19(7), 683-687. <https://doi.org/10.1080/13504851.2011.595673>
- Warner, R. M. (1998). *Spectral analysis of time-series data*. Guilford Press.
- Woo, J., & Owen, A. L. (2019). Forecasting private consumption with Google Trends data. *Journal of Forecasting*, 38(2), 81-91. <https://doi.org/10.1002/for.2559>

World Bank. (2021). Malaysia Economic Monitor, June 2021: Weathering the Surge.
<https://openknowledge.worldbank.org/handle/10986/35812>

World Bank. (2022, November 29). *The World Bank in Malaysia: Overview*. Retrieved 2024, January 11 from <https://www.worldbank.org/en/country/malaysia/overview>