Nowcasting Cost of Living Using Google Trends: The Case of Selected Asian Countries

Yee Sheng Mak[1] | *School of Mathematical Sciences, Universiti Sains Malaysia, Minden, Penang, Malaysia*
Siok Kun Sek[2] | *School of Mathematical Sciences, Universiti Sains Malaysia, Minden, Penang, Malaysia*
Mohsen Ayyash[3] | *School of Mathematical Sciences, Universiti Sains Malaysia, Minden, Penang, Malaysia*

Received 14.10.2025 (revisions received 13.3.2026, 30.3.2026), **Accepted** (reviewed) 5.4.2026, **Published** 11.9.2026

Abstract

This paper examines the cost of living (COL) in six selected Asian countries: Indonesia, Japan, Korea, Malaysia, Singapore, and Thailand. Motivated by rising inflationary pressures and limitations of traditional economic indicators – such as reporting delays and limited granularity – this study explores whether real-time data from Google Trends searches can improve the tracking and prediction of COL patterns in several country-specific settings. The analysis covers the period from 2010 to 2024 and employs Mixed Data Sampling (MIDAS) regression. The results demonstrate that MIDAS models provide superior flexibility and responsiveness by accommodating high-frequency Google Trends data alongside lower-frequency macroeconomic variables. Notably, MIDAS models yield improved predictive accuracy and timeliness in capturing COL dynamics. The findings also highlight the significance of search terms like debt, inflation, and unemployment as behavioural proxies for public economic concerns. This study underscores the potential of integrating search engine data into COL analysis to improve the timeliness and sensitivity of economic monitoring.

Keywords	***DOI***	***JEL code***
Cost of living, Google Trends, macroeconomic indicators, Mixed Data Sampling (MIDAS) regression, real time data	*https://doi.org/10.54694/stat.2025.56*	*C82, C22, R20, C53, E31, D12, D31*

INTRODUCTION

The cost of living (COL) denotes the resources required to sustain a standard of living, based on essential expenditures such as housing, food, transport, and healthcare. It is a key measure of household welfare

[1] School of Mathematical Sciences, Universiti Sains Malaysia, 118 00 USM Penang, Malaysia.
[2] School of Mathematical Sciences, Universiti Sains Malaysia, 118 00 USM Penang, Malaysia. Corresponding author: e-mail: sksek@usm.my. <ORCID ID: https://orcid.org/0000-0002-3616-9815>.
[3] School of Mathematical Sciences, Universiti Sains Malaysia, 118 00 USM Penang, Malaysia. <ORCID: https://orcid.org/0000-0002-5618-6313>.

and an important input into wage and policy decisions (ILO, 2022; OECD, 2011). In Asia, COL has risen sharply in recent decades, shaped by rapid urbanisation, structural transformation, and persistent inflation. These pressures have been exacerbated by external shocks: the 2008 Global Financial Crisis triggered food and fuel price surges that hit low-income households hardest, while the COVID-19 pandemic caused severe income losses and drove an estimated 90 million people into extreme poverty worldwide, many in Asia (FAO, 2021).

Asia's diversity makes it a compelling setting for COL analysis. High-income economies such as Singapore, Japan, and South Korea consistently rank among the world's most expensive countries, driven by housing, urban costs, and currency strength. By contrast, South and Central Asian nations such as Pakistan, Bangladesh, and India remain among the least expensive globally, reflecting lower wages, rural consumption patterns, and modest price levels (Clissitt, 2023). Southeast Asia lies between: Malaysia and Thailand record mid-range living costs but show steeper upward trends compared to advanced peers. This heterogeneity underscores how development stages and economic structures shape COL dynamics.

Traditional COL measures, including the consumer price index (CPI) and household surveys, remain central but face well-known shortcomings: delayed publication, lack of regional granularity, and reliance on static consumption baskets that fail to capture behavioural shifts during shocks (Cavallo, 2013; ILO, 2022; OECD, 2011). These limitations highlight the value of complementary, higher-frequency indicators. Digital data offers such potential. Real-time information sources enhance responsiveness and allow for timelier detection of economic changes (Polash, 2026). Google Trends, in particular, provides a cost-free measure of public search activity and has been widely used for nowcasting unemployment, inflation, and consumer sentiment (Choi and Varian, 2012; D'amuri and Marcucci, 2017; Vosen and Schmidt, 2011). Its applications have broadened significantly across disciplines (Jun et al., 2018), though concerns remain about reliability for low-frequency terms (Gummer and Oehrlein, 2024). Despite this, its potential for monitoring COL specifically remains underexplored, especially in emerging Asian contexts. To date, no study has applied Google Trends search data in analyzing the cost of living. Hence, utilizing the real-time data using Google Trends search in studying COL could contribute to new insights in understanding the behaviour of households, forecasting analysis, and policy planning or decision-making.

This study investigates COL dynamics in six Asian economies – Indonesia, Japan, South Korea, Malaysia, Singapore, and Thailand – between 2010 and 2024. By comparing Malaysia with both regional peers and advanced economies, the analysis highlights differences in COL trajectories across development levels. The study pursues three objectives: (i) to evaluate the relationship between macroeconomic indicators and COL, (ii) to assess the explanatory power of Google Trends data, individually and as a composite index, and (iii) to test whether hybrid models integrating digital and traditional data, particularly through MIDAS methods, yield improved insights into COL dynamics.

1 BACKGROUND STUDY – GOOGLE TRENDS

Google Trends is a tool that provides data on the relative popularity of search queries over time. It enables users to explore patterns in search behaviour, with data normalized so that the highest search volume in a selected period is assigned a value of 100, and all other values are scaled accordingly. This tool is widely used in economic research for forecasting key indicators, analyzing consumer behaviour, and predicting market movements (Choi and Varian, 2012). For instance, researchers have employed Google Trends data to anticipate changes in unemployment rates, housing activity, and even financial markets.

Google Trends offers a powerful example of real-time data in practice. As a freely accessible platform, it provides continuous updates on search activity, reflecting the interests, concerns, and behaviours of populations around the world. Its applications span across business, policymaking, and research. For businesses, Google Trends is invaluable for market research, consumer insights, and forecasting.

It helps identify seasonal demand patterns, detect emerging product categories, and benchmark brand visibility against competitors. In the field of digital marketing, it guides search engine optimization (SEO) strategies by revealing rising keywords and enabling micro-targeted content creation (Polash, 2026). Geographically, Google Trends allows for precise regional analysis of search activity, supporting localized marketing campaigns and enabling researchers to uncover cultural or regional differences in public concerns (Hölzl, 2025). From a policy perspective, it has been increasingly used to nowcast economic indicators such as unemployment, inflation, and consumer sentiment, offering governments a low-cost supplement to traditional surveys, particularly in data-scarce environments (Hölzl, 2025). Its role in crisis detection and public health monitoring has also been well documented, with search spikes serving as early warning signals for events such as influenza outbreaks and the COVID-19 pandemic. In academia and journalism, Google Trends contributes to the study of human behaviour, opinion dynamics, and media coverage by providing a rich source of digital behavioural data (Hölzl, 2025). Importantly, it is not only versatile but also cost-effective, offering free, continuously updated insights that can be integrated into decision-making at multiple levels.

The practical applications of these insights are wide-ranging. In economics, Google Trends data has been incorporated into forecasting models for unemployment, inflation, and housing market dynamics, allowing policymakers and analysts to generate faster estimates than traditional surveys permit. In public health, search patterns have been used to detect outbreaks of influenza and, more recently, to track public concern and behavioural responses during the COVID-19 pandemic. In business, companies use real-time analytics and Google Trends to monitor consumer demand, identify emerging product categories, and adapt their marketing strategies to shifting public interests. These applications illustrate the versatility of Google Trends as both a predictive and diagnostic tool, particularly in contexts where conventional data sources are delayed, costly, or incomplete.

Beyond these practical uses, methodological innovation represents another crucial area of application. The integration of Google Trends with other big data sources, such as social media feeds, economic indicators, or mobility data, has opened new avenues for interdisciplinary research. This approach not only mitigates the limitations of search-only data but also enhances the robustness of findings by triangulating across multiple streams of real-time information. Studies such as those by Jun et al. (2018) and Gummer & Oehrlein (2024) collectively demonstrate that while Google Trends is an accessible and cost-efficient source of data, its full potential is realized when applied with careful attention to methodological considerations and combined with complementary datasets.

While Google Trends is a powerful, free tool for analyzing search behaviour and identifying public interest, it is not a direct measure of absolute search volume. Users must be cautious in utilizing and interpreting the information. Google Trends data are not a raw data, but it is a normalized, relative data (Jun et al., 2018). Such data is useful for comparing and analyzing changes in trends over time, but it cannot provide precise and absolute measurements. In addition, online search might be reliable to reflect personal situation, but it is less reliable when the search variables are unknown to the individual (Niesert et al., 2020). Google Trends search is also claimed to lack sentiment. It shows the volume of search for a topic, but it does not tell why or what sentiment the search has. High frequency data, such as daily data, is only available for short periods. Google Trends cannot provide for long-term trend analysis (over several years) without losing precision.

There are many cases of improper application and analysis using data from Google Trends by researchers. Hölzl et al. (2025) conducted a systematic review on the misuse of Google Trends applications in the social sciences. They revealed many improper sampling and estimation procedures or mistakes by researchers in the literature. The majority of studies did not perform an internal validity test, did not concern about data reliability, and did not concern about the generalizability of results. Failing to address these issues leads to misleading and inaccurate results collected from Google Trends.

Together, these applications underline the dual nature of Google Trends as both an opportunity and a challenge. Its broad accessibility and near real-time availability make it a powerful instrument for academia, business, and public policy. The literature demonstrates that when applied appropriately, Google Trends offers significant value as a tool for forecasting, behavioural analysis, and decision-making in fast-changing environments (Gummer and Oehrlein, 2024; Jun et al., 2018).

2 LITERATURE REVIEW

2.1 Theoretical Review

The cost of living is underpinned by several core economic theories that collectively explain how prices, consumption, and policy interact over time. Index number theory provides the statistical foundation for measuring changes in economic variables, particularly prices, through the construction of index numbers relative to a base period. Central to this is the cost of living index, which quantifies the expenditure required to maintain a constant standard of living over time. The concept originated with Konüs (1939) and was advanced by Frisch (1936). Traditional price indices such as the Laspeyres and Paasche indices, have notable limitations in accounting for shifts in consumer behaviour when relative prices change (Triplett, 1992). To address these limitations, superlative indices, such as the Fisher Ideal and Törnqvist indices were developed. Theoretical justification for these was provided by Diewert (1983), who demonstrated that such indices offer accurate approximations of the "true" cost-of-living index. These indices are widely used for inflation measurement and cost of living analysis due to their ability to reflect substitution effects and real-world price dynamics.

Complementing this, consumer choice theory explains how individuals allocate limited resources to maximize utility amidst changing prices and income. Two key mechanisms are the substitution effect and the income effect (Keat and Young, 2009; Latimaha et al., 2018). Rising living costs tend to restrict substitution, especially for inelastic necessities such as healthcare and housing. This theory highlights how price changes disproportionately affect lower-income households, emphasizing the importance of public policy in mitigating inflation's impact (Latimaha et al., 2018).

Adding a global dimension, trade openness theory explores how liberalization of trade and capital flows influences domestic prices and consumer welfare. Greater openness can reduce inflationary pressures through access to cheaper imports and productivity gains (Latimaha et al., 2018; Squalli and Wilson, 2006). However, it also increases exposure to external shocks such as global commodity price fluctuations and exchange rate volatility. For open economies like Malaysia, currency depreciation raises import costs, thereby exacerbating household financial burdens (Lafrance and Schembri, 2000; Latimaha et al., 2018). While trade openness can enhance consumer welfare, it necessitates policy responses to manage distributional impacts and external vulnerabilities.

Finally, macroeconomic theory provides a broader context for understanding COL dynamics. The Phillips Curve illustrates the trade-off between inflation and unemployment, with both influencing household welfare (Cebula & Toma, 2008; Squalli & Wilson, 2006). Monetary and fiscal policies – including interest rate adjustments and government spending – play crucial roles in controlling inflation and shaping household affordability (International Monetary Fund, 2015; Latimaha et al., 2018; United States Congress, 1996). Wage dynamics are also pivotal; while rising wages can help preserve real incomes, they may contribute to inflation via a wage–price spiral (Blanciforti and Kranner, 1997; Cebula, 1980). Effective policymaking requires balancing growth, price stability, and protection of real incomes, especially for households vulnerable to economic shocks.

2.2 Empirical Review

Consumer Price Index (CPI), representing overall inflation, consistently shows a positive relationship with the cost of living. Bosire and Omwenga (2022) found that inflation increased household living expenses in Kajiado North, Kenya. Similarly, Najaf and Najaf (2016) observed that oil price volatility and exchange rate fluctuations heightened the cost of living in Pakistan through inflationary effects. Latimaha et al. (2018) reported similar findings for Malaysia, identifying inflation as a key driver of rising living costs. Overall, CPI emerges as a consistently significant and positive contributor to living costs across diverse countries and methodologies.

Gross Domestic Product (GDP), commonly used as a proxy for national income or economic growth, typically influences the cost of living through rising income levels and increased demand for goods and services. Jenkins (2000) found that temporary increases in household income in the UK led to short-term rises in living costs. Cebula and Todd (2004) similarly reported a positive association between economic activity – linked to GDP – and the cost of living in Florida counties. Although many studies focus on income or wages rather than GDP directly, the evidence consistently supports a positive relationship between GDP and cost of living.

Population size, especially in urban areas, is also positively associated with living costs. Haworth and Rasmussen (1973) observed that larger U.S. metropolitan populations incurred higher living expenses due to increased demand for housing and public services. Cebula and Toma (2008) confirmed these findings, noting that population density was positively correlated with the cost of living in U.S. counties. In Malaysia, Latimaha et al. (2018) identified urbanization as a significant factor contributing to elevated living costs. Collectively, these studies indicate that larger and denser populations drive up living expenses, particularly in urban settings.

Exchange rate fluctuations, especially currency depreciation, are widely found to increase the cost of living by raising the price of imported goods. Hobdari and Uppal (2017) showed that depreciation led to food price shocks and higher living costs in Sub-Saharan Africa. Similarly, Najaf and Najaf (2016) found that exchange rate instability contributed to inflation and increased living expenses in Pakistan. Lafrance and Schembri (2000) confirmed this relationship in Canada, where currency depreciation was linked to higher living costs. Overall, the literature consistently identifies a positive relationship between exchange rate depreciation and the cost of living.

Housing and property costs are frequently identified as key contributors to living expenses. Kurre (2000) reported that housing prices significantly increased the cost of living in Pennsylvania counties. Cebula (1980) and Cebula and Toma (2008) also found that housing costs and urban amenities were positively associated with living costs. In Malaysia, Latimaha et al. (2018) highlighted housing expenses as a major determinant of living costs in capital cities. Additionally, Bridel and Lontoh (2014) showed that the removal of fuel subsidies raised transportation costs, which could indirectly impact housing affordability through commuting expenses. Across all reviewed studies, housing and property prices are consistently linked to higher living costs.

Unemployment shows a more complex and indirect relationship with the cost of living. Latimaha et al. (2019) found that unemployment in Malaysian states indirectly contributed to higher living costs by exacerbating urban crime issues. However, unemployment is not commonly analyzed as a direct determinant of living costs. The limited evidence suggests a positive but indirect link, indicating a need for further research in this area.

2.3 Research Gaps

A thorough review of existing literature reveals significant gaps in methodology, data usage, and analytical scope that this study seeks to address. A critical shortcoming in existing research is its reliance on conventional macroeconomic indicators – such as inflation, GDP growth, and employment – that are typically delayed in reporting, aggregated at national levels, and insufficient for capturing individuals' and communities' lived economic realities. These top-down measures lack the granularity required to reflect public sentiment and the day-to-day experiences of economic pressure, thereby limiting their relevance for timely and localized policy decisions. As economies face increasingly rapid shifts, especially during crises, the need for real-time, high-frequency data has grown. Google Trends, by reflecting what people search for in real time, offers an immediate and cost-effective complement to traditional statistics.

Previous studies also have some shortcomings in terms of methodology, scope, and data used. A major methodological limitation is the predominance of conventional linear regression models, such as ordinary least squares (OLS). For instance, studies by Blanciforti and Kranner (1997) and Cebula and Todd (2004) used linear techniques to analyse cost-of-living trends in the U.S., often neglecting complex economic behaviours such as nonlinearities, structural changes, and asymmetric responses. These approaches may oversimplify interactions and fail to capture the nuanced behaviour of cost-related variables across different countries and time periods.

Geographical and temporal constraints further weaken the generalizability of prior research. Many studies are confined to specific regions or policy events – for example, Bridel and Lontoh's (2014) study of Malaysia's 2013 fuel reform or Bosire and Omwenga's (2022) work in a Kenyan sub-county. Such narrow scopes hinder comparative and longitudinal analysis. Next, data limitations also constrain the utility of existing studies. Most rely on low-frequency, aggregated macroeconomic indicators from official sources. While reliable, these data suffer from reporting lags, lack household-level detail, and fail to capture real-time public sentiment. This disconnect between delayed statistics and immediate lived experience weakens the policy relevance of cost-of-living research. As a result, there is a growing recognition of the importance of real-time data sources that can complement and enhance traditional measures. In recent years, Google Trends has gained prominence in economic research, particularly in forecasting unemployment, consumer behaviour, and financial markets. However, its application to COL analysis remains limited, leaving an important gap that this study seeks to fill.

To overcome these issues, researchers have begun exploring advanced methodologies and novel data sources. Mixed Data Sampling (MIDAS) models offer a robust framework for integrating high- and low-frequency data without sacrificing information richness. Rather than relying on macroeconomic indicators, which are slow and delay in capturing economic behaviours, this study applies a new approach by integrating traditional macroeconomic indicators with alternative digital data – specifically, Google Trends search behaviour. Real-time search queries related to inflation, housing, healthcare, and government subsidies are analyzed to construct thematic indices that act as proxies for public attention and sentiment. These digital indicators complement official statistics by capturing emerging concerns as they unfold, allowing for earlier detection of economic stress compared to lagged macroeconomic measures.

3 DATA AND METHODOLOGY

3.1 Data Collection and Processing

This study is focused on six selected Asian countries: Indonesia (IDN), Japan (JAP), Korea (KOR), Malaysia (MYS), Singapore (SGP), and Thailand (THA). These countries represent two main clusters, the developed Asia (JAP, KOR, and SGP) versus the developing Asia (IDN, MYS, and THA). This study seeks to compare the cost of living (COL) in these two clusters in Asia, reflecting the COL situation of very high COL versus those that are experiencing increasing COL in Asia. The results might not fully represent the conditions of all countries in Asia, but they could tell the conditions of COL in many countries in Asia.

The data are in different frequencies, covering the range of 2010 to 2024. This study uses COL as the main variable of interest, measured annually using data from Numbeo, a global database that compiles user-reported consumer prices and living costs. Numbeo, one of the most comprehensive databases of cost-of-living data, calculates its index based on crowdsourced prices for core goods and services such as food, rent, utilities, and transportation. The index is benchmarked against New York, which is assigned a value of 100. For example, Kuala Lumpur, Malaysia, has a cost of living index of 33.3, implying that the overall expenses there are approximately 33.3% of those in New York. An index value below 100 indicates lower relative costs, while a value above 100 reflects higher expenses. This index plays a key role in wage benchmarking, relocation planning, business site selection, and policy formulation. Nonetheless, its accuracy may be affected by inconsistencies in data collection, currency volatility, and differing consumption patterns across regions.

The data collection process of Numbeo involves user-generated input and manually gathered information from reputable sources. The reliance on such crowdsourced data can be prone to human error and seasonal fluctuations (Numbeo, 2026). The cost coverage also overlooks local, non-consumer costs such as taxes, healthcare, and childcare, and it does not account for individual lifestyle differences. The index is calculated from a fixed „basket of goods" for a hypothetical four-person family (Numbeo, 2026). Such a rigid lifestyle weighting fails to account for individual spending habits. Furthermore, the use of New York City as a baseline may not well reflect the consumption habits of residents in other locations. Despite these limitations, Numbeo offers a platform and free access to its extensive dataset for COL analyses and comparisons, covering thousands of cities, providing various specific metrics for international comparisons

In addition to COL, several macroeconomic indicators were included to capture broader economic and price dynamics: consumer price index (CPI), gross domestic product (GDP), population size (POP), real effective exchange rate (RER), real residential property price index (RPI), and unemployment rate (URT). These variables reflect key dimensions of economic performance, labour market conditions, inflationary pressure, and housing affordability. Data for these indicators were obtained from multiple sources, including Numbeo, the International Monetary Fund (IMF), the Federal Reserve Economic Data (FRED), the Department of Statistics Malaysia (DOSM), and the Singapore Department of Statistics (SINGSTAT).

Table 1 summarizes the variables used, including their descriptions, frequencies, sources, and relevant notes. For GDP, a natural logarithm transformation was applied to account for its large scale; the transformed variable is referred to as LNGDP in the analysis. Variables reported annually, namely COL and POP, were converted to quarterly frequency using linear interpolation. Linear interpolation is a mathematical technique used to estimate unknown values that fall between two known data points by assuming a straight line exists between them. It provides simple and easy implementation. It is very effective when the data follows a smooth trend as linear interpolation can consistently yield the lowest error rates and perform better than much more complex machine learning models (Zuhairi et al., 2025).

Table 1 List of variables used in the analysis

Variable	Description	Frequency	Source	Remarks
COL	Cost of living index	Annual	Numbeo	
CPI	Consumer price index	Monthly	IMF	IDN data obtained from FRED
GDP	Gross domestic product	Quarterly	IMF	MYS data from DOSM; log- transformed as LNGDP
POP	Population (millions)	Annual	IMF	
RER	Real effective exchange rate	Monthly	FRED	
RPI	Real residential property price index	Quarterly	FRED	
URT	Unemployment rate (%)	Quarterly	IMF	MYS data from DOSM; SGP data from SINGSTAT

Source: Authors' computation

Complementing macroeconomic indicators, the study incorporates monthly Google Trends data for six search terms linked to economic concerns: *debt*, *inflation*, *loan*, *recession*, *tax*, *and unemployment*. These search indices, retrieved in monthly frequency without aggregation, serve as proxies for public sentiment and offer high-frequency insights into cost-of-living pressures.

The following four model specifications were estimated:

Model 1: COL= F (CPI, LNGDP, POP, RER, RPI, URT)

Model 2: COL= F (*debt*, *inflation*, *loan*, *recession*, *tax*, *unemployment*)

Model 3: COL= F(PC1)

Model 4: COL= F (CPI, LNGDP, POP, RER, RPI, URT, PC1)

In the above, PC1 refers to the first principal component obtained from Principal Component Analysis (PCA) of the six Google Trends indices. The difference between Model 2 and Model 3 is, Model 2 is based on keywords search using Google Trends, each keyword is associated with a single search volume index. Then, in Model 3, all six indices from Model 2 are integrated to generate a composite index using PCA method. As a summary, Model 1 is based on macroeconomic data alone, Model 2 and Model 3 are regressed on Google Trends data while Model 4 consists of macroeconomic data and Google Trends data. Each model was estimated using two approaches.

3.2 Methodology – MIDAS

Mixed Data Sampling (MIDAS) regression is an econometric technique designed to incorporate regressors observed at higher frequencies into a regression with a lower-frequency dependent variable, without resorting to temporal aggregation. The method allows for flexible and parsimonious modeling of lag effects from high-frequency predictors by using weighting schemes that restrict or parameterize the lag structure.

The general form of the MIDAS regression is:

$$y_t = \mathbf{X}_t^{(L)'}\beta + f\left(\{\mathbf{X}_t^{(H)}\} : \nu, \ell\right) + \varepsilon_t$$

where y_t is the dependent variable observed at low frequency, $\mathbf{X}_t^{(L)}$ is the vector of low-frequency regressors with coefficients β; $\mathbf{X}_t^{(H)}$ is a set of high-frequency regressors, $f(\cdot)$ is a lag weighting function parameterized by ν over ℓ high-frequency lags, and ε_t is the error term.

Several forms of MIDAS regressions differ in how they specify the weighting function for the high-frequency lags. In the step-weighted approach, the high-frequency lags are grouped into steps of equal length, with each step sharing the same coefficient. Another approach uses polynomial distributed lag (PDL) weighting, also known as Almon lags, which constrains the lag coefficients to lie on a polynomial curve. This method yields smoothly varying weights that can capture gradual decay, rise, or hump-shaped lag effects.

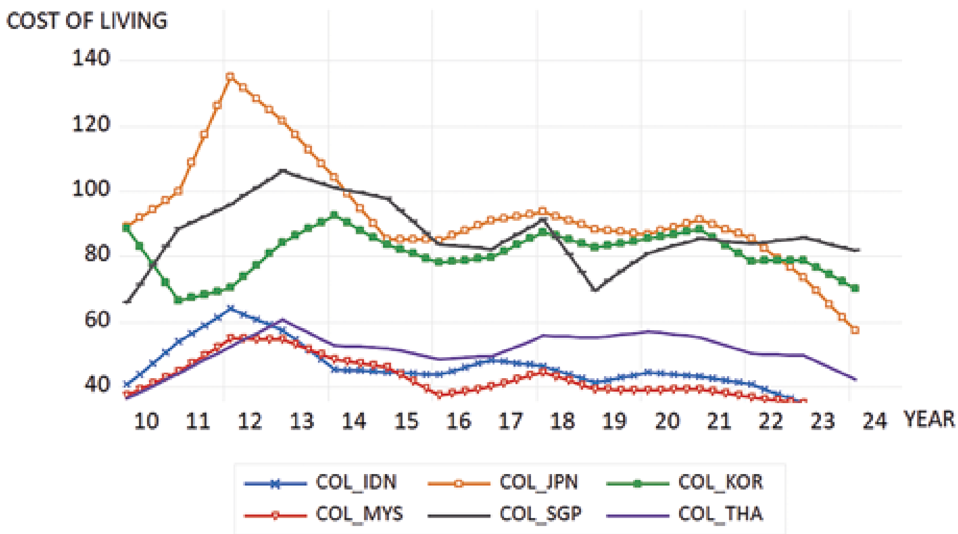
In contrast to these structured models, the unrestricted MIDAS (U-MIDAS) formulation allows each high-frequency lag to enter the model with its own freely estimated coefficient. To improve parsimony while retaining the flexibility of U-MIDAS, the general-to-specific (GETS-MIDAS) procedure begins with the full set of unrestricted lags and applies a model selection algorithm to eliminate those that are statistically insignificant. The selection is typically based on information criteria such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), resulting in a sparser model that retains only the most informative lags.

4 RESULTS AND DISCUSSION

This section presents the key empirical findings of the study, focusing on the comparative performance of different estimation models and techniques in modelling and predicting the cost of living (COL) across selected Asian countries. The primary aim is to evaluate whether incorporating Google Trends data enhances model performance relative to relying solely on traditional macroeconomic variables.

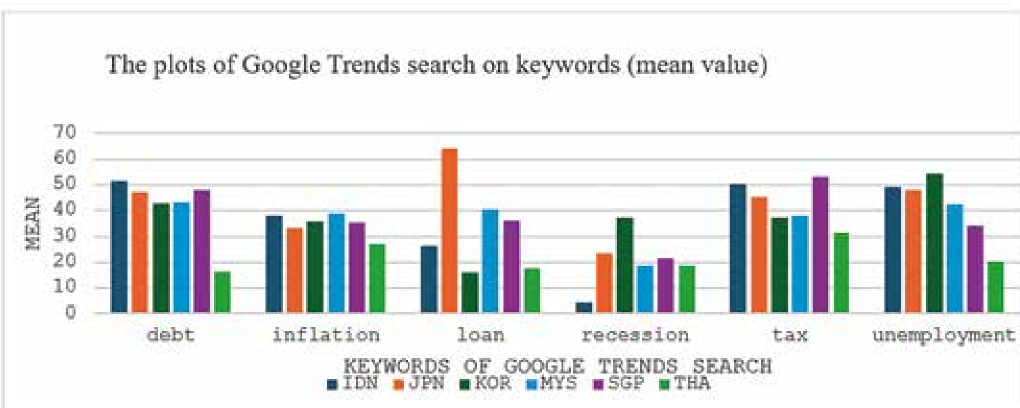
Below are the plots of COL and regressors (macroeconomics versus Google Trends keywords indices).

Figure 1 Cost of living (2010–2024)



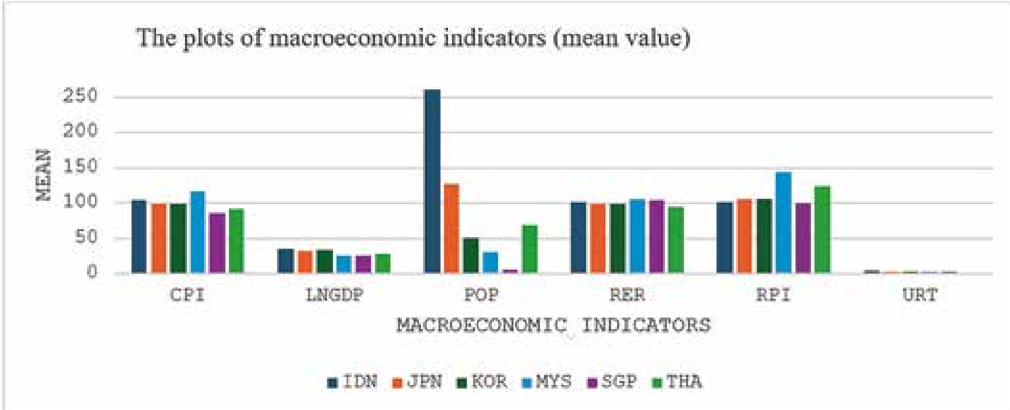
Source: Authors' computation

Figure 2 Mean values of Google Trends keywords indices



Source: Authors' computation

Figure 3 Mean values of macroeconomic indicators



Source: Authors' computation

The cost of living plots (Figure 1) show that there are two clusters of COL where Japan, Korea, and Singapore are under the same cluster with higher COL; Indonesia, Malaysia, and Thailand are in the lower COL cluster. The cluster with higher COL includes countries with higher economic development compared to the lower cluster with lower economic development. All countries show a declining trend in the 2020s relative to New York, U.S.

In terms of Google Trends search (Figure 2), each keyword search is accompanied by a core index. These indices reflect the main concern of individuals. For instance, Japan is more concerned about loans; while Singapore and Indonesia are concerned about taxes and debt, and Korea is concerned about unemployment. Malaysia is more concerned about inflation. In general, keywords debt, tax, and unemployment show higher values; these keywords are searched more frequently, while the word recession has a much lower search volume.

In terms of macroeconomic indicators (Figure 3), Malaysia exhibited higher mean values (2010–2024) in inflation, real exchange rate, and real property index than other countries. Japan and Indonesia are leading in terms of population size. Thailand also showed a relatively higher mean in the real property index. These macroeconomic factors might contribute to higher COL in respective countries, but they may not track the behaviour of households in time.

4.1 Model Performance by Country

The analysis involves the following steps. First of all, preliminary tests (unit-root and cointegration) are performed to examine the properties of the data. Unit-root tests reveal that all variables are not stationary at the level but are stationary in first differences. As all variables are stationary in first differences, the cointegration test is performed and reveals the existence of a long-run relationship. Hence, all models (Model 1, 2, 3, and 4) are estimated at the level form (long-run relationship) by comparing MIDAS (PDL-MIDAS, STEP-MIDAS, U-MIDAS, and GETS-MIDAS) with the benchmark OLS. Results are evaluated using five metrics. Finally, forecast averaging analysis is performed on the best model to proxy COL for each country.

Model evaluation is based on standard metrics: Adjusted R-squared, Loglikelihood, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). Higher Adjusted R-squared and less negative Loglikelihood values indicate better model fit, while lower RMSE and MAE suggest improved forecast precision. Across the four specified models, estimations were conducted using Ordinary Least Squares

(OLS) and four variants of Mixed Data Sampling (MIDAS) regression – PDL-MIDAS, STEP-MIDAS, U-MIDAS, and GETS-MIDAS.

Tables 2 to 5 present the detailed evaluation metrics for Models 1 to 4, respectively, across all six countries in the study. Each table reports the values of Adjusted R-squared, Loglikelihood, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) for the five estimation methods: OLS, PDL-MIDAS, STEP-MIDAS, U-MIDAS, and GETS-MIDAS. These metrics allow for a comprehensive comparison of model fit and predictive accuracy, highlighting the strengths and weaknesses of each estimation approach in different country-model combinations. For each country and metric, the best-performing value is highlighted in bold and italics to facilitate identification of the most accurate method. Additionally, the final column in each table indicates the best overall model for each country, based on a holistic evaluation of all metrics.

Table 2 Evaluation metrics for Model 1

Country	Method	R-SQ	LOGLIKE	RMSE	MAE	Best Model
IDN	OLS	***0.9872***	***–60.2265***	0.7705	***0.6323***	OLS
	PDL-MIDAS	0.9824	–67.1751	0.8594	0.6925	
	STEP-MIDAS	0.9826	–68.0871	0.8743	0.7296	
	U-MIDAS	0.9837	–60.9400	***0.7640***	0.6336	
	GETS-MIDAS	0.9845	–64.4392	0.8162	0.6578	
JPN	OLS	***0.9833***	–112.6004	2.0251	***1.4550***	OLS
	PDL-MIDAS	0.9799	–116.8464	2.1062	1.5416	
	STEP-MIDAS	0.9800	–117.9158	2.1483	1.5637	
	U-MIDAS	0.9797	–112.9325	***1.9589***	1.4696	
	GETS-MIDAS	0.9818	***–99.8482***	11.6083	8.2500	
KOR	OLS	0.9182	–101.3504	1.6378	1.3984	U-MIDAS
	PDL-MIDAS	0.9229	–99.8482	1.5374	1.2535	
	STEP-MIDAS	0.9208	–101.8469	1.5954	1.2863	
	U-MIDAS	0.9030	***–92.9654***	***1.3534***	***1.0785***	
	GETS-MIDAS	***0.9329***	–96.1009	1.7897	1.4996	
MYS	OLS	0.9815	–64.4726	0.8167	0.6698	U-MIDAS
	PDL-MIDAS	0.9819	–60.9549	0.7482	0.6103	
	STEP-MIDAS	0.9828	–61.4998	0.7557	0.6065	
	U-MIDAS	0.9873	***–44.5608***	***0.5523***	***0.4312***	
	GETS-MIDAS	***0.9901***	–46.0247	0.5674	0.4388	
SGP	OLS	0.9365	–112.6106	2.0254	1.4560	U-MIDAS
	PDL-MIDAS	0.9391	–110.9186	1.8872	1.4027	
	STEP-MIDAS	0.9407	–111.4155	1.9047	1.4240	
	U-MIDAS	0.9317	***–109.8249***	***1.8494***	***1.3992***	
	GETS-MIDAS	***0.9424***	–112.4420	1.9412	1.4481	
THA	OLS	0.9347	–71.5218	0.9329	0.7742	U-MIDAS
	PDL-MIDAS	0.9470	–67.4159	0.8432	0.7012	
	STEP-MIDAS	0.9467	–68.8133	0.8654	0.7142	
	U-MIDAS	***0.9487***	***–62.3800***	***0.7682***	***0.6191***	
	GETS-MIDAS	0.9468	–70.5475	0.8936	0.6775	

Source: Authors' computation

Table 3 Evaluation metrics for Model 2

Country	Method	R-SQ	LOGLIKE	RMSE	MAE	Best Model
IDN	OLS	***0.9958***	–106.5034	4.7127	3.4624	U-MIDAS
	PDL-MIDAS	0.9697	–78.9397	1.0438	0.7963	
	STEP-MIDAS	0.9716	–81.5466	1.0955	0.8692	
	U-MIDAS	0.9810	***–45.9353***	***0.5665***	***0.4494***	
	GETS-MIDAS	0.9864	–62.0820	15.5234	13.9179	
JPN	OLS	***0.9969***	–214.8034	7.4933	5.6438	U-MIDAS
	PDL-MIDAS	0.9784	–117.8481	2.0481	1.6732	
	STEP-MIDAS	0.9793	–119.4305	2.1113	1.7425	
	U-MIDAS	0.9901	***–73.3924***	***1.0637***	***0.8632***	
	GETS-MIDAS	0.9833	–140.7901	2.7096	2.2567	
KOR	OLS	***0.9972***	–154.9581	5.2479	4.0543	U-MIDAS
	PDL-MIDAS	0.9803	–83.1075	1.2652	1.0338	
	STEP-MIDAS	0.9806	–83.8267	1.2908	1.0549	
	U-MIDAS	0.9910	***–48.5823***	***0.6561***	***0.5195***	
	GETS-MIDAS	0.9893	–69.1165	1.0057	0.8807	
MYS	OLS	***0.9911***	–125.3293	6.2892	4.4566	U-MIDAS
	PDL-MIDAS	0.9700	–84.2465	2.2122	1.8016	
	STEP-MIDAS	0.9713	–85.6197	2.2753	1.8533	
	U-MIDAS	0.9807	***–57.3927***	***1.2815***	***1.0614***	
	GETS-MIDAS	0.9787	–70.0708	1.6547	1.4036	
SGP	OLS	***0.9963***	–137.4049	4.6192	3.5945	U-MIDAS
	PDL-MIDAS	0.9744	–73.9705	1.0722	0.8770	
	STEP-MIDAS	0.9744	–73.9803	1.0726	0.8773	
	U-MIDAS	0.9870	***–40.7327***	***0.5705***	***0.4581***	
	GETS-MIDAS	0.9833	–56.0417	0.8561	0.6995	
THA	OLS	***0.9986***	–115.1474	2.8051	2.1862	U-MIDAS
	PDL-MIDAS	0.9807	–67.6515	0.9093	0.7461	
	STEP-MIDAS	0.9809	–67.9927	0.9207	0.7549	
	U-MIDAS	0.9891	***–30.9165***	***0.4290***	***0.2847***	
	GETS-MIDAS	0.9879	–54.2076	0.8252	0.6733	

Source: Authors' computation

Table 4 Evaluation metrics for Model 3

Country	Method	R-SQ	LOGLIKE	RMSE	MAE	Best Model
IDN	OLS	***0.9900***	***−58.6022***	***0.7311***	***0.4845***	OLS
	PDL-MIDAS	0.9587	−97.1553	1.4626	1.0704	
	STEP-MIDAS	0.9596	−97.0758	1.4605	1.0720	
	U-MIDAS	0.9568	−96.6511	1.4490	1.0785	
	GETS-MIDAS	0.9608	−97.2928	1.4664	1.0602	
JPN	OLS	0.9588	−139.2548	3.3485	2.6539	U-MIDAS
	PDL-MIDAS	0.9579	−140.9368	3.2904	2.5678	
	STEP-MIDAS	0.9588	−140.9193	3.2893	2.5714	
	U-MIDAS	***0.9851***	***−110.5077***	***1.8729***	***1.0235***	
	GETS-MIDAS	0.9597	−140.8451	3.2848	2.5856	
KOR	OLS	***0.9145***	***−105.3153***	1.7650	1.4938	U-MIDAS
	PDL-MIDAS	0.9056	−109.4711	1.8373	1.5563	
	STEP-MIDAS	0.9079	−109.3660	1.8337	1.5455	
	U-MIDAS	0.9080	−107.0721	***1.7575***	***1.4601***	
	GETS-MIDAS	0.9085	−109.7051	1.8453	1.5547	
MYS	OLS	***0.9658***	−83.6015	1.1717	0.9583	U-MIDAS
	PDL-MIDAS	0.9718	−78.4188	1.0338	0.8567	
	STEP-MIDAS	0.9721	−78.6740	1.0387	0.8679	
	U-MIDAS	0.9706	***−77.7849***	***1.0218***	***0.8342***	
	GETS-MIDAS	0.9711	−80.1721	1.0679	0.8568	
SGP	OLS	***0.9196***	***−121.6273***	***2.4011***	***1.8327***	OLS
	PDL-MIDAS	0.9098	−125.6759	2.4803	1.8731	
	STEP-MIDAS	0.9114	−125.7198	2.4824	1.8703	
	U-MIDAS	0.9074	−124.6798	2.4350	1.8495	
	GETS-MIDAS	0.9148	−125.7297	2.4828	1.8646	
THA	OLS	0.9143	***−81.5154***	***1.1265***	***0.9433***	OLS
	PDL-MIDAS	0.9134	−84.8369	1.1643	1.0210	
	STEP-MIDAS	0.9155	−84.7220	1.1618	1.0150	
	U-MIDAS	0.9130	−83.2799	1.1312	0.9525	
	GETS-MIDAS	***0.9169***	−84.8085	1.1637	1.0056	

Source: Authors' computation

Table 5 Evaluation metrics for Model 4

Country	Method	R-SQ	LOGLIKE	RMSE	MAE	Best Model
IDN	OLS	***0.9883***	–57.3356	0.7288	0.6101	U-MIDAS
	PDL-MIDAS	0.9832	–63.9516	0.8087	0.6446	
	STEP-MIDAS	0.9840	–64.7096	0.8204	0.6634	
	U-MIDAS	0.9841	***–55.2170***	***0.6858***	***0.5626***	
	GETS-MIDAS	0.9850	–62.8451	13.1594	10.8055	
JPN	OLS	***0.9839***	–111.0176	1.9655	***1.3930***	U-MIDAS
	PDL-MIDAS	0.9798	–114.9708	2.0343	1.5244	
	STEP-MIDAS	0.9810	–115.3348	2.0481	1.5403	
	U-MIDAS	0.9783	***–109.7732***	***1.8476***	1.3938	
	GETS-MIDAS	0.9812	–110.8418	94.5502	93.3076	
KOR	OLS	0.9232	–99.0843	1.5692	1.3386	U-MIDAS
	PDL-MIDAS	0.9441	–89.1627	1.2614	1.0180	
	STEP-MIDAS	0.9466	–89.9252	1.2793	1.0397	
	U-MIDAS	0.9390	***–84.4279***	***1.1555***	***0.9634***	
	GETS-MIDAS	***0.9488***	–90.6846	1.2975	1.0498	
MYS	OLS	0.9822	–62.9381	0.7934	0.6383	U-MIDAS
	PDL-MIDAS	0.9845	–54.6518	0.6657	0.5435	
	STEP-MIDAS	0.9846	–57.1992	0.6979	0.5679	
	U-MIDAS	0.9896	***–33.2582***	***0.4480***	***0.3506***	
	GETS-MIDAS	**0.9910**	–42.6560	0.5331	0.4257	
SGP	OLS	0.9365	–111.9823	2.0016	1.4639	U-MIDAS
	PDL-MIDAS	0.9388	–109.0439	1.8228	1.3988	
	STEP-MIDAS	***0.9418***	–109.6673	1.8440	1.4379	
	U-MIDAS	0.9333	***–104.2763***	***1.6688***	***1.2850***	
	GETS-MIDAS	0.9374	–115.2292	2.0441	1.4498	
THA	OLS	0.9376	–69.6867	0.9011	0.7450	U-MIDAS
	PDL-MIDAS	0.9443	–66.7720	0.8333	0.6880	
	STEP-MIDAS	0.9451	–68.4000	0.8588	0.6945	
	U-MIDAS	0.9441	***–59.7833***	***0.7321***	***0.5644***	
	GETS-MIDAS	***0.9468***	–79.5475	0.8936	0.6775	

Source: Authors' computation

In Model 1, which includes only macroeconomic variables, performance differed by country, with both OLS and U-MIDAS achieving strong results. In Indonesia and Japan, OLS outperformed the MIDAS variants, particularly in terms of Adjusted R-squared and MAE, suggesting that in these contexts, the macroeconomic indicators captured most of the relevant variation in COL. However, in Korea, Malaysia, Singapore, and Thailand, U-MIDAS consistently outshone OLS, delivering lower RMSE and higher loglikelihood values. This pattern suggests that in these countries, even for macroeconomic data, the flexible structure of U-MIDAS allowed for better exploitation of temporal patterns. Notably, GETS- MIDAS occasionally recorded high R-squared values but was penalized by its higher forecast errors, confirming that good in-sample fit does not necessarily translate to better predictive performance.

Model 2, which relies exclusively on individual Google Trends indices, demonstrated the strongest predictive capacity overall. U-MIDAS was the clear leader in every country, achieving the best loglikelihood, RMSE, and MAE. Although OLS reported slightly higher R-squared values in some cases, the difference was marginal and did not compensate for its inferior forecast accuracy. The superiority of U-MIDAS in this model highlights the model's ability to exploit the real-time nature of search trends data, which may reflect emerging concerns about the cost of living more promptly than traditional indicators. PDL-MIDAS and STEP-MIDAS were moderately effective but failed to challenge U-MIDAS in any country. GETS-MIDAS again underperformed, particularly in Indonesia, where its forecast errors were significantly higher.

For Model 3, which utilises a composite Google Trends index generated via Principal Component Analysis (PCA), performance varied more widely. In Indonesia and Singapore, OLS outperformed all MIDAS methods, possibly due to the strong explanatory power of the composite index in these contexts. Conversely, U-MIDAS was the top performer in Korea and Malaysia, achieving the lowest RMSE and MAE, suggesting that in these cases, the flexible dynamic weighting of search data within U-MIDAS was beneficial. Japan presented a mixed outcome: while OLS had a higher R-squared, U-MIDAS achieved the best RMSE and loglikelihood, tipping the balance in favour of U-MIDAS. Thailand's results were relatively balanced, though OLS narrowly led in most performance metrics. The mixed findings for Model 3 suggest that a composite index may dilute specific signals present in individual search series, resulting in performance differences depending on country-specific dynamics.

In Model 4, which combines macroeconomic variables with the composite Google Trends index, U- MIDAS emerged as the most robust and reliable method in five out of six countries. It consistently achieved the lowest RMSE and highest loglikelihood, reflecting strong predictive ability and good model fit. Japan was the only exception, where OLS surpassed U-MIDAS in R-squared and MAE, although U-MIDAS still had better RMSE and loglikelihood. PDL-MIDAS, STEP-MIDAS, and GETS-MIDAS occasionally ranked highly in specific metrics but could not match U-MIDAS's overall consistency. These results underscore the added value of integrating behavioural data with economic indicators, particularly when processed through a flexible modelling framework such as U-MIDAS.

4.2 Comparative Insights

Table 6 reports the estimation performance metrics for all four models across the six selected countries. The metrics include R-squared (R-SQ), log-likelihood (LOGLIKE), root mean square error (RMSE), and mean absolute error (MAE). For each country, the best value for each metric is highlighted in bold and italics. The best overall model for each country – determined by the collective performance across all metrics – is indicated in the final column. This comparative analysis provides a comprehensive evaluation of model fit and forecasting accuracy across different model specifications and countries.

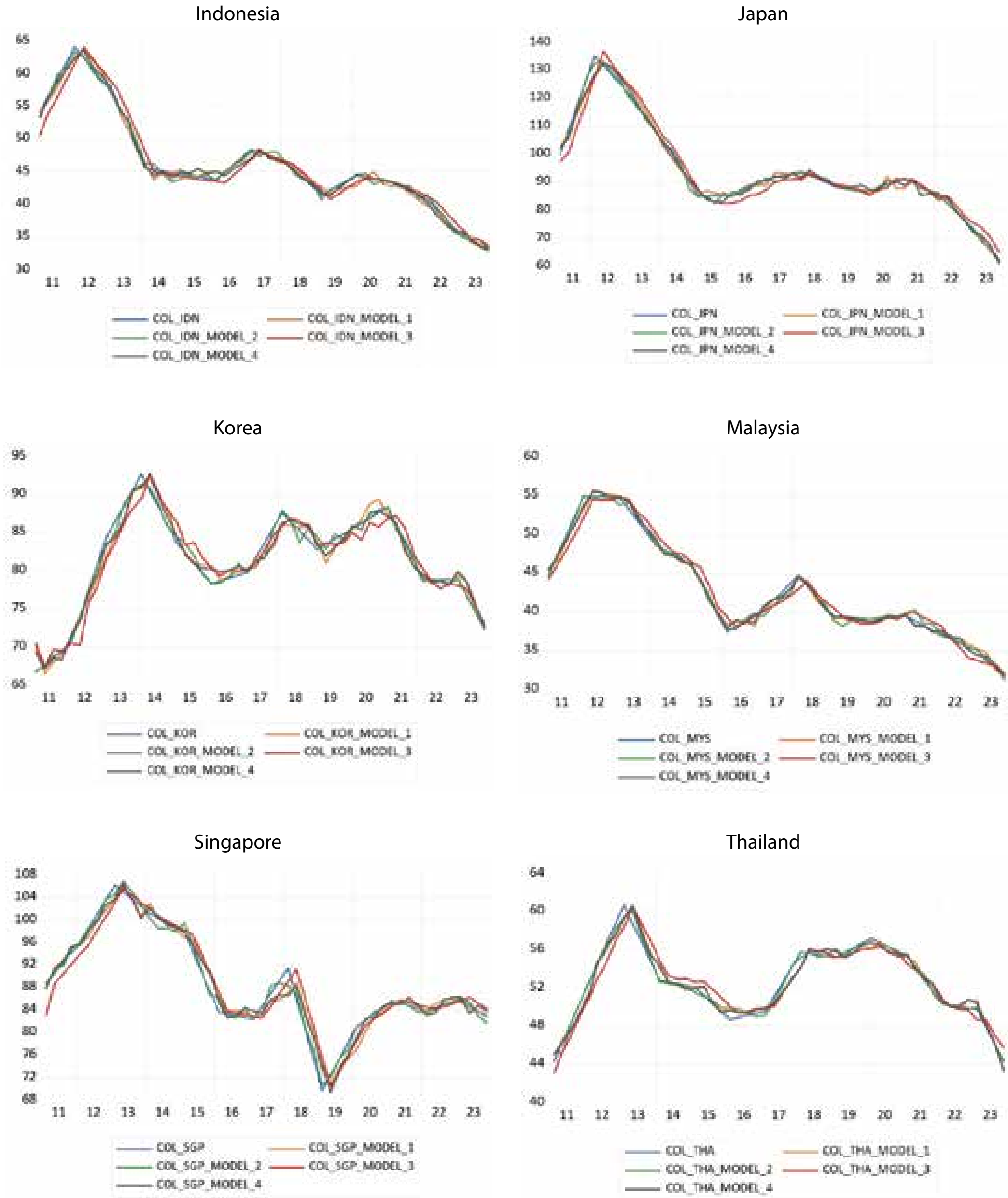
Table 6 Estimation Performance Metrics and Best Model Selection

Country	Model	R-SQ	LOGLIKE	RMSE	MAPE	Best Model
IDN	1	***0.9872***	-60.2265	0.7705	1.3803	2
	2	0.9810	***-45.9353***	***0.5665***	***0.9672***	
	3	0.9900	-58.6022	0.7311	1.0074	
	4	0.9841	-55.2170	0.6858	1.2178	
JPN	1	0.9833	-112.6004	2.0251	1.4507	2
	2	***0.9851***	***-82.6423***	***1.1179***	***0.9515***	
	3	0.9851	-110.5077	1.8729	1.0555	
	4	0.9783	-109.7732	1.8476	1.4610	
KOR	1	0.9030	-92.9654	1.3534	1.3312	2
	2	***0.9569***	***-58.0789***	***0.7094***	***0.5946***	
	3	0.9080	-107.0721	1.7575	1.7959	
	4	0.9390	-84.4279	1.1555	1.1908	
MYS	1	0.9873	-44.5608	0.5523	0.9951	4
	2	0.9812	-37.1707	0.4816	0.8535	
	3	0.9706	-77.7849	1.0218	1.9427	
	4	***0.9896***	***-33.2582***	***0.4480***	***0.8144***	
SGP	1	0.9317	-109.8249	1.8494	1.6331	2
	2	***0.9532***	***-78.2403***	***1.0304***	***0.8819***	
	3	0.9074	-124.6798	2.4350	2.1621	
	4	0.9333	-104.2763	1.6688	1.5077	
THA	1	0.9487	-62.3800	0.7682	1.1739	2
	2	***0.9891***	***-30.9165***	***0.4290***	***0.5426***	
	3	0.9143	-81.5154	1.1265	1.8053	
	4	0.9441	-59.7833	0.7321	1.0685	

Source: Authors' computation

Figure 4 provides the plots of real COL data (in dark blue line) versus the estimated COL under different models. In most cases, Model 2 outperforms the other models as COL estimated using Model 2 is the closest to the real COL data.

Figure 4 Estimated COL versus real COL data



Source: Authors' computation

The comparative analysis yields several important insights. First, models that incorporate Google Trends data – especially Model 2 and Model 4 – generally outperform the baseline macroeconomic model (Model 1). Specifically, Model 2 was the best-performing specification in five out of six countries, while Model 4 led only in Malaysia, where Model 2 was a close second. This suggests that search data, either on its own or in combination with traditional indicators, provides a richer and more timely signal for modelling the cost of living. Model 3, which uses a composite index, was the weakest performer across all countries, reinforcing the idea that disaggregated search data retains more useful information than a single aggregated measure.

Secondly, the results emphasize the importance of using appropriate estimation techniques. MIDAS models – particularly U-MIDAS – consistently outperformed OLS in terms of forecast accuracy, regardless of the model specification. This confirms the methodological advantage of MIDAS in handling mixed-frequency data, allowing for better exploitation of real-time search trends.

Finally, these findings have broader implications. They demonstrate that integrating behavioural data, such as Google search trends, with traditional economic analysis can enhance the monitoring and forecasting of COL pressures. This is relevant for both developed and developing economies, as policymakers seek more responsive tools to detect and manage economic challenges. The effectiveness of MIDAS models in this context highlights their value in real-time economic analytics, supporting data-driven policy interventions that are timely and adaptive.

4.3 Discussion

Google Trends keywords search reveals that people from different countries show their specific concerns, reflecting their cost of living particulars. Relatively, Japan is highly concerned about 'loan', Singapore is more concerned about 'debt' and 'tax', Korea is relatively more concerned about 'debt', 'recession' and 'unemployment'; Indonesia is concerned about 'tax' and 'unemployment', while Malaysia is concerned about 'inflation'. These keywords are the main search and could predict the movement of COL better than the macroeconomic factors, as they reflect the real-time needs and behaviour shifts. These keywords searches can be classified into financial stress (such as debt, loan), economic uncertainty or shock (recession, inflation), and policy stance (tax). These factors are recognized to be influential in nowcasting COL across countries, but with different magnitudes. Comparing the results between the developed versus developing clusters, there is no clear-cut on which factor may dominate COL according to the country development cluster.

In general, loans, debt, and taxes are more relevant or influential, while recession is less influential. These three determinants are more influential as they could induce higher financial burden and risk, lower purchasing power, and lower living standard. As discussed in Chai et al. (2020), indebtedness may expose financial stress and financial risk, and other economic implications, including a reduction of future consumption. On the other hand, the impose of taxes on goods and services induces higher prices. An increase in one key product triggers a cyclical chain reaction of price increases in many more products across the economy, leading to cost-push inflation. The increase of income tax causes a lower spending ability. All these lead to higher COL. According to Maina et al. (2024), Direct taxation is more influential than indirect taxation in affecting consumer burden and the house price index. The loan also affects COL in the same way by boosting higher financial stress and lowering spending ability. As COL is defined as the amount needed to sustain a minimal living standard, the increase of debt, loans, and taxes will lead to a drop in money available to sustain the same living standard, leading to a spike in COL.

The level of development does not directly determine the factors leading to the increase of COL, but country-specific factors (e.g., economic structure, economic stability, changes of policies) and living environment/styles might have more direct impacts on COL. Hence, the results might not fully represent the two clusters within Asia or outside Asia. Rather, it reflects the relative influence of each factor across countries by comparing countries with different characteristics.

Next, the nowcasting results show that all six Asian countries are expected to show a lower cost of living compared to the benchmark of the U.S., with Korea and Thailand show deeper declining trend than the others and Singapore might remain high in COL. The results provide useful information to the policymaker to focus on the keywords and macroeconomic variables that show greater magnitude in each country, as these are the main determinants of COL. Such integration of macroeconomic data and Google Trends data could be utilized by policymakers in other countries in analyzing the dynamics of COL.

CONCLUSIONS

This study investigated the potential of Google Trends data as a real-time, high-frequency indicator for modelling and forecasting the cost of living (COL) in six Asian countries: Indonesia, Japan, Korea, Malaysia, Singapore, and Thailand. Traditional macroeconomic data often suffer from reporting delays and limited granularity. In contrast, search engine data provide a timely and novel approach to capturing public interest and sentiment related to economic conditions. Four model specifications were estimated using both Ordinary Least Squares (OLS) and several MIDAS regression techniques. These models incorporated macroeconomic variables, Google Trends data, or a combination of both. Across all countries, MIDAS regression methods outperformed OLS, underscoring the advantages of mixed-frequency modelling for tracking dynamic economic phenomena.

Key findings indicate that Model 2 – based solely on Google Trends indices – achieved the highest predictive accuracy in Indonesia, Japan, Korea, Singapore, and Thailand. In Malaysia, Model 4, which combines macroeconomic indicators with a composite Google index, performed best; however, Model 2 was a close second. This highlights the strong predictive value of search behaviour data even in countries with moderate digital infrastructure. Furthermore, individual Google search indices proved more informative than aggregated indices, reinforcing the usefulness of specific search terms for economic modelling.

These findings carry important policy implications. The integration of real-time digital behavioural data with traditional macroeconomic indicators can greatly enhance the responsiveness and precision of cost-of-living monitoring systems. Policymakers, particularly in developing and digitally engaged economies, should consider incorporating search engine data to complement official statistics. This can support more timely and targeted interventions during inflationary periods or economic shocks. Researchers could also explore the use of other digital data sources, such as social media trends or e-commerce data, and apply alternative modelling approaches, including machine learning techniques. Additionally, investigating the causal mechanisms behind search behaviour and its relationship with economic indicators would deepen the understanding of how digital activity reflects real-world conditions.

Although Google Trends is a powerful tool for data search behaviour, it also has some downsides. Issues of reliability, particularly for low-frequency search terms, necessitate careful validation and methodological refinement. Researchers should pay caution in utilizing and interpreting the information. In the era of data analytics and big data, data is valuable. Google Trends is a very useful tool if it is applied with careful attention to methodological considerations and combined with complementary datasets.

While Google Trends is highly utilized to make data-driven decisions, including anticipating shifts in consumer demand/market trends, optimizing marketing strategies, and content planning, its applications in policy decision and planning are lacking behind. The policymakers, governmental organizations and research institutions could apply Google Trends to monitor public interest/reactions, nowcast economic trends, and observe real-time and fast-moving trends to provide better forecasts on economic scenarios. This tool is also useful to detect the geological disasters (the outbreak of diseases, floods condition, etc.) and provide early warning on the change of symptoms or hotspot detection. In terms of decision-making, the real-time data is feasible in making forecasts, revealing fast-moving trends to gauge public concerns and responses, detecting outbreaks of diseases, so that immediate action can be taken in emergency situations. It also provides additional information in policy-decision, cost-effective planning and development.

In conclusion, this study provides compelling evidence that Google Trends data can enhance the modelling and forecasting of cost-of-living dynamics. The successful integration of real-time behavioural data with traditional economic indicators offers valuable tools for researchers and policymakers to improve economic analysis, surveillance, and policy response in an increasingly digitalised world.

DECLARATION OF GENERATIVE AI IN THE WRITING

We did not apply any AI tool in the writing process.

References

BLANCIFORTI, L. A., KRANNER, E. (1997). *Estimating county cost of living indexes: The issue of urban versus rural.* [online]. Regional Research Institute Working Paper No. 9718, National Institute for Occupational Safety and Health, Morgantown. [cit. 20.8.2025]. <https://researchrepository.wvu.edu/rri_pubs/186/>.

BOSIRE, W. M., OMWENGA, J. Q. (2022). Determinants of cost of living of the households in Kajiado North Sub-County. *International Journal of Social Science and Humanities Research*, 10(4): 148–167. <https://doi.org/10.5281/zenodo.7183190>.

BRIDEL, A., LONTOH, L. (2014). *Lessons Learned: Malaysia's 2013 Fuel Subsidy Reform* [online]. Geneva: International Institute for Sustainable Development. [cit. 20.8.2025].<https://www.iisd.org/sites/default/files/gsi/ffs_malaysia_lessonslearned.pdf>

CAVALLO, A. (2013). Online and official price indexes: Measuring Argentina's inflation. *Journal of Monetary Economics*, 60(2): 152–165. < https://doi.org/10.1016/j.jmoneco.2012.10.002>.

CEBULA, R. J. (1980). Determinants of geographic living-cost differentials in the United States: An empirical note. *Land Economics*, 56(4): 477–481. < https://doi.org/10.2307/3146225>.

CEBULA, R. J., TODD, S. (2004). An empirical note on determinants of geographic living-cost differentials for counties in the state of Florida, 2003. *The Review of Regional Studies*, 34(1): 112–119. <https://doi.org/10.52324/001c.8372.>.

CEBULA, R. J., TOMA, M. (2008). An empirical analysis of determinants of interstate living-cost differentials, 2005. *Journal of Regional Analysis and Policy*, 38(3): 1–7. <https://doi.org/ 10.22004/ag.econ.133000>.

CHAI, Y., ABDULLAH YUSOF, S., MOHD AMIN, R., MOHD ARSHAD, M. N. (2020). Household Debt and Household Spending Behaviour: Evidence from Malaysia. *Jurnal Ekonomi Malaysia*, 54(1): 111–120. <http://dx.doi.org/10.17576/JEM-2020-5401-8>.

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>

CLISSITT, C. (2023). *Cost of living around the world in 2023* [online]. London: MoveHub. [cit. 20.8.2025]. <https://www.movehub.com/advice/cost-of-living-around-the-world/>.

D'AMURI, F., MARCUCCI, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, 33(4): 801–816. <https://doi.org/10.1016/j.ijforecast.2017.03.004>.

DEPARTMENT OF STATISTICS MALAYSIA (DOSM). (2024). *Gross domestic product and unemployment rate statistics* [online]. Putrajaya: Department of Statistics Malaysia. [cit. 20.8.2025]. <https://open.dosm.gov.my/dashboard>.

DIEWERT, W. E. (1983). The theory of the cost-of-living index and the measurement of welfare change. In: DIEWERT, W. E., MONTMARQUETTE, C (eds.) *Price Level Measurement*. Ottawa: Statistics Canada.

FEDERAL RESERVE BANK OF ST. LOUIS. (2024). *FRED economic data* [online]. St. Louis: Federal Reserve Bank of St. Louis. [cit. 20.8.2025]. < https://fred.stlouisfed.org/>.

FOOD AND AGRICULTURE ORGANIZATION (FAO). (2021). FAO Statistical Yearbook 2021 [online]. Rome: FAO. [cit. 20.8.2025]. <https://openknowledge.fao.org/server/api/core/bitstreams/522c9fe3-0fe2-47ea-8aac-f85bb6507776/content>.

FRISCH, R. (1936). The problem of index numbers. *Econometrica*, 4: 1–38. <https://doi.org/10.2307/1907119>.

GUMMER, T., OEHRLEIN, A.-S. (2024). Using Google Trends data to study high-frequency search terms: Evidence for a reliability-frequency continuum. *Social Science Computer Review*, 43(4): 814–826. <https://doi.org/10.1177/08944393241279421>.

HAWORTH, C. T., RASMUSSEN, D. W. (1973). Determinants of metropolitan cost of living variations. *Southern Economic Journal*, 40(2): 183–192. <https://doi.org/10.2307/1056869>.

HOBDARI, M. N. A., UPPAL, A. (2017). *Food Inflation in Sub-Saharan Africa: Causes and Policy Implications* [online]. International Monetary Fund. [cit. 20.8.2025]. <https://www.imf.org/external/pubs/ft/wp/2016/wp16247.pdf>.

HÖLZL, J., KEUSCH, F., SAJONS, C. (2025). The (mis)use of Google Trends data in the social sciences - A systematic review, critique, and recommendations. *Social science research,* 126, 103099. <https://doi.org/10.1016/j.ssresearch.2024.103099>.

INTERNATIONAL MONETARY FUND (IMF). (2015). *Malaysia: 2015 Article IV consultation – staff report*. Washington, D.C.: International Monetary Fund. [cit. 20.8.2025]. <https://www.elibrary.imf.org/view/journals/002/2015/058/002.2015.issue-058-en.xml>.

INTERNATIONAL LABOUR ORGANIZATION (ILO). (2022). *Understanding Cost of Living Indicators* [online]. Geneva: ILO. [cit. 20.8.2025]. <https://share.google/gJ9Pf8r2KLtX72KNd>.

JENKINS, S. P. (2000). Modelling household income dynamics. *Journal of Population Economics*, 13(4): 529–567. <https://doi.org/10.1007/s001480050151>.

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>.

KEAT, P. G., YOUNG, P. K. Y. (2009). *Managerial Economics: Economic Tools for Today's Decision Makers*. Pearson Education.

KONÜS, A. A. (1939). The problem of the true index of the cost of living. *Econometrica*, 7(1): 10–29. <https://doi.org/10.2307/1906997>.

KURRE, J. A. (2000). *Differences in the Cost of Living Across Pennsylvania's 67 Counties*. Harrisburg: The Center for Rural Pennsylvania. [cit. 20.8.2025]. <https://www.rural.pa.gov>.

LAFRANCE, R., SCHEMBRI, L. L. (2000). *The exchange rate, productivity, and the standard of living*. [online]. Bank of Canada Review. [cit. 20.8.2025]. <https://www.bankofcanada.ca/wp-content/uploads/2010/06/r001b-e.pdf>.

LATIMAHA, R., BAHARI, Z., ISMAIL, N. A. (2018). A descriptive analysis on the basic needs budget by the middle income earners to identify the most expensive city to live: An analysis involving capital cities in Malaysia. *Malaysian Journal of Business and Economics*, 5(2): 1–18. <https://doi.org/10.51200/mjbe.v0i0.1606>.

LATIMAHA, R., BAHARI, Z., ISMAIL, N. A. (2019). Examining the linkages between street crime and selected state economic variables in Malaysia: A panel data analysis. *Malaysian Journal of Economics*, 53(1): 59–72. <http://dx.doi.org/10.17576/JEM-2019-5301-6>.

MAINAL, S. A., ABU HANIPAH, F. A., ZUHAIMI, N. L., KAMARUDIN, N. N. S. (2024). The Impact of Taxation on the Cost of Living: A Comprehensive Analysis. *Information Management and Business Review*, 16(3(I)S), 628–643. <https://doi.org/10.22610/imbr.v16i3(I)S.4089N>.

NUMBEO. (2024). *Cost of Living Index Database* [online]. Belgrade: Numbeo. [cit. 20.8.2025]. <https://www.numbeo.com/cost-of-living/>.

NAJAF, R., NAJAF, K. (2016). Impact of oil prices on the stock exchange of Pakistan and Malaysia. *International Journal of Research* -GRANTHAALAYAH, 4(8): 84-91. <https://doi.org/10.5281/zenodo.61161>.

NUMBEO. (2026). *Methodology and motivation*. [online]. Belgrade: Numbeo. [cit. 20.8.2025]. <https://www.numbeo.com/common/motivation_and_methodology.jsp>.

NIESERT, R. F., OORSCHOT, J. A., VELDHUISEN, C. P., BRONS, K., LANGE, R. J. (2020). Can Google search data help predict macroeconomic series? *International Journal of Forecasting*, 36(3): 1163–1172. <https://doi.org/10.1016/j.ijforecast.2018.12.006.>.

POLASH, A. (2026). *How to use Google Trends for research and stay ahead of your competitors*. [online]. Florida: Hosting.com. [cit. 20.8.2025]. <https://hosting.com/blog/how-to-use-google-trends-for-research/>.

SINGAPORE DEPARTMENT OF STATISTICS (SINGSTAT). (2024). *Unemployment Rate Statistics* [online]. Singapore: Singapore Department of Statistics. [cit. 20.8.2025]. < https://www.singstat.gov.sg/>.

SQUALLI, J., WILSON, K. (2006). A new approach to measuring trade openness. Economic and Policy Research Unit Working Paper 06–07, Zayed University, Dubai, UAE. <https://www.zu.ac.ae/epru/documents/06-07-web.pdf>.

TRIPLETT, J. E. (1992). Economic theory and BEA's alternative quantity and price indexes. *Survey of Current Business*, 72: 49–52.

UNITED STATES CONGRESS. (1996). Toward a more accurate measure of the cost of living: final report to the Senate Finance Committee. Dec. 4, 1996 updated version. [Washington, D.C.]: Advisory Commission to Study the Consumer Price Index. <https://catalog.hathitrust.org/Record/003239902>.

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>.

ZUHAIRI, A. H., YAKUB, F., HARUM, A. N., OMAR, M., SHARIFUDDIN, M., FARUQ, A., SINHA, V., RAZAK, K. A., & ABDULLAH, S. (2025). Comparative analysis of missing data imputation methods for flood features from Langat river in Selangor, Malaysia. *Malaysian Journal of Computer Science*, 38(3): 224–237. <https://doi.org/10.22452/mjcs.vol38no3.2>.