STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–17.
Published online in International Academic Press (www.IAPress.org)

Statistical methods for inflation forecasting in Morocco: Insights from Google trends data

Mariem Bikourne [1], Sokaina El khamlichi [2,*], Adil Ez-Zetouni [1], Khadija Akdim [1]

[1] *Department of Mathematics, Faculty of Sciences and Technology, Cadi Ayyad University, 549, 40.000, Marrakesh, Morocco*
[2] *LyRICA – Laboratory of Research in Computer Science, Data Sciences and Knowledge Engineering, School of Information Sciences, Rabat, Morocco*

Abstract
Accurate inflation forecasting is essential for effective economic planning and policy-making. The increasing use of the internet enables user generated content to capture people's expectations and perspectives on economic issues. This study aims to investigate the power of Google trends data as an effective complementary source of data for forecasting inflation in Morocco. By identifying keywords that exhibit Granger causality with the inflation rate, we examined the linear effect of public interest on inflation forecasting using a principal component index as an exogenous factor to enhance outcomes. The selected SARIMA model, coupled with the resulting index, presents an optimal trajectory for inflation rate. The results of this study demonstrate that the model incorporating Google Trends data yielded the best performance based on evaluation measures such as AIC, RMSE, and log-likelihood. This highlights that the Google index is a significant factor for accurately explaining and forecasting inflation rate movements, contributing substantially to inflation modeling. The adaptive features of our approach make it preferably suited to describing inflation uncertainty when the economy is subject to constantly changing monetary institutions and policies.

Keywords Inflation Rate, Forecasting, SARIMAX, Inflation Uncertainty, Google Trends, Principal Component Analysis, Web Mining

DOI: 10.19139/soic-2310-5070-2172

1. Introduction

The consumer price index (CPI) is a metric that measures the average change in the prices paid by a typical consumer for a common basket of goods and services. The CPI estimates the purchase power of one unit of currency to determine the average cost of living in a specific country. As such, it is the primary macroeconomic indicator for gauging inflation (or deflation) and enhancing the precision of inflation projections [1]. Consequently, the CPI influences many different market dynamics and is an essential driver of economic activity [2]. In the mid-1980s, developed countries experienced the Great Moderation, which was a period of significant disinflation. During this period, there was consistent low inflation and moderate economic development [3]. Later, the Global Financial Crisis (GFC) of 2008, and more recently, the consequences of the Covid-19 outbreak on the world economy [4], prompted extraordinary monetary interventions, potentially affecting global inflation dynamics [5] [6] [7]. Economists acknowledge

*Correspondence to: Sokaina El khamlichi (Email: sokaina.elkhamlichi@gmail.com). Laboratory of Research in Computer Science, Data Sciences and Knowledge Engineering, School of Information Sciences, Rabat, Morocco

ISSN 2310-5070 (online) ISSN 2311-004X (print)

the significance of current inflation research, measures, and estimation, despite ongoing debates on the fundamental causes of inflation [8].

The ability to effectively foresee the forthcoming inflation rate is highly valued by policymakers and market participants. Indeed, inflation forecasting is a crucial instrument for altering monetary policies across the world [9][10]. Central banks forecast future inflation patterns in order to justify interest rate choices as well as control and keep inflation around the selected target [11]. An improved knowledge about the upcoming inflation trends can assist decision-makers in formulating effective monetary policy [12]. Inflation forecasting is particularly crucial for fiscal authorities that want to alter social security payments and support packages for certain economic sectors. In the private sector, fixed income investors want to forecast future inflation in order to predict forthcoming patterns in discounted real returns. Furthermore, certain private companies must anticipate particular components of inflation in order to foresee pricing trends and adjust for emerging risks. Lastly, the anticipated trajectory of inflation has a significant impact on the amounts of debt held by the public and private sectors as well as interest payments [13].

Some forecasting techniques, such as time series approaches, econometric techniques, and intelligent algorithms, have been used to forecast inflation. Traditional and extensively utilized time series and econometric techniques include autoregressive integrated moving average (ARIMA) family [14] [15] [16], fractional modeling approach [17], Error Correction Model (ECM) [18], Autoregressive Distributed Lag (ADL) [19], Exponential Smoothing approach [20], non-linear dimension reduction techniques [21], Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models [22][23] [24], autoregressive models with time varying parameters [25], Vector autoregression Models (VAR) [10], and dynamic factor models [26] [27]. Moreover, there are several studies that exploits forecast combination methods for forecasting inflation. For instance, Silva et al. [28] forecast the food Consumer Price Index, and inflation employing the parametric and nonparametric approaches of Exponential Smoothing, Holt-Winters, ARIMA, as well as Singular Spectrum Analysis. On the other hand, intelligent algorithms such as support vector machines [29], neural networks [30, 31, 8], and genetic algorithms [32] are also employed to predict economic indicators. Nevertheless, such economic metrics are typically issued at regular intervals and after the occurrence of actual economic crises. For instance, the CPI for June is accessible in mid-July. It indicates that statistical information lags about fifteen days after the publication of economic indicators. In the past, this period was not long enough to impede the expansion of the economy.

Conversely, in the information era, lags may contribute to economic downturn if inadequate policies are made during this period. The conventional source of data is then distinguished by its low frequency and emission delays. To enhance inflation forecasting accuracy, novel sources of data with frequent availability and no lags are urgently required. Thereby, it is necessary to provide new forecasting techniques incorporating new data sources at various frequencies [33]. Actually, forecasting economic indicators using web searches data are recently implemented. For instance, Bouayad et al. investigated the power of Google Trends data as an effective substitute source of data for forecasting and nowcasting Gross Domestic Product growth in Morocco [34]. Furthermore, great attempts are made to anticipate inflation in the short and medium term. However, not much emphasis is placed on how accurate inflation nowcasts are. According to Faust and Wright [3], firmly establishing inflation nowcasts is essential for producing reliable inflation predictions, as major inaccuracies in the nowcast spread across the whole horizon. Moreover, Monteforte and Moretti [35] use a mixed-frequency model to analyze inflation in the eurozone and conclude that, in comparison to models that rely on monthly data, the inclusion of daily variables reduces nowcast errors. Modugno [36] shows that the accuracy of the nowcast is increased by inputting high-frequency data using a trading day frequency factor approach. In contrast to the traditional studies that use large datasets, Knotek and Zaman [37] examine a limited number of data series at different frequencies in order to provide an explanation for the nowcast. They assert that the model is generally more accurate than the Blue Chip consensus nowcasts of headline CPI or the Survey of Professional Forecasters, and that it yields historically lower RMSFEs than MIDAS regression, similar to Monteforte and Moretti [35] and the DFM model found on [36].

According to the World Bank Data website [38], 88% of Moroccan population use the internet to find information. As a result, with the widespread use of internet searches, we now own more information about our beliefs, expectations, and identities than ever before. About ten years ago, it would have been impossible to obtain information on human behaviour from these data. Internet search activity might be taken as an indicator of disclosed expectations. People are likely to seek information on topics about which they desire to learn more or on issues which concern them. For instance, if a person is concerned about managing home spending and feels prices are growing, they may seek information about inflation. Indeed, consumers may be concerned about inflation in an environment where wages are less volatile than prices. Moreover, if Google search queries for "inflation" are increasing, it may indicate popular concern about rising costs. The amount of inflation-related search inquiries indicates the public's expectations for future price adjustments.

Early uses of social media and internet search query metadata analysis were in the fields of epidemiology and diseases symptom surveillance [39, 40]. Furthermore, attempts to model financial and economic systems in the context of human behavior have spanned recent studies such as stock market movements [41], unemployment [42], GDP [43], oil prices [44], inflation and consumption [45]. The first attempt to forecast inflation using Google Trends data comes from the work of [46]. The author compares and introduces a new measure of inflation predictions based on Google search queries. However, a small number of researches have been implemented to forecast inflation using GT data.

In this work, we investigate the effectiveness of Google trends (GT) data as a complementary source of data for forecasting inflation in Morocco. We use Principal Component Analysis (PCA) to reduce dimensionality in GT queries and construct an explanatory variable of the selected model. More specifically, this study contributes to existing inflation forecasting models by addressing seasonality in the Moroccan inflation rate. We employ univariate models to analyze the inflation time series, incorporating Seasonal Autoregressive (SAR) and Seasonal Moving Average (SMA) terms into higher-order Autoregressive Integrated Moving Average (ARIMA) models. This approach was implemented using the SARIMA model to forecast inflation, and subsequently, a SARIMAX model was applied to incorporate the PCA index as an exogenous variable.

The primary objective of this study is to produce accurate monthly inflation forecasts in Morocco by leveraging Google Trends data and to determine if a statistically significant relationship exists between Google Trends data and the inflation rate in real time. Additionally, the study evaluates the added value of Google Trends as a complementary data source for inflation forecasting.

The remainder of this paper is organized as follows. Section 2 presents the data and methodology of this work. Section 3 deals with the results and discussion. Section 4 sums up the present investigation.

2. Data and Methodology

In this section, we discuss data used in our application and provide preliminary results for the keywords selection. Next, we introduce our modeling framework and the main theoretical methods used for the formulated problem. Figure 3 analytically summarizes and illustrates the steps of our methodology.

2.1. Data Description

2.1.1. Consumer Price Index data: The dataset used in this study is collected from the International Monetary Fund (IMF) [47]. We extracted monthly data for Morocco, spanning from January 2018 to September 2023. This period was selected for this analysis because this timeframe encompasses recent economic cycles that captures both stable periods and significant economic shifts, including the impacts of the COVID-19 pandemic and subsequent global economic disruptions. These events have had substantial effects on inflation, supply chains, and consumer prices, providing a robust basis for examining inflationary trends and their underlying drivers

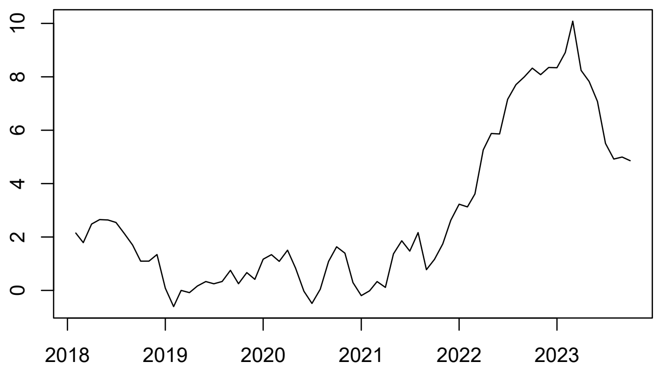

Figure 1. Inflation time series data for the Moroccan economy between January 2018 and September 2023

The Consumer Price Index (CPI) was used to measure the (YoY%) inflation rate as follow

$$\mathrm{Inf}_t = \frac{\mathrm{CPI}_t - \mathrm{CPI}_{t-1}}{\mathrm{CPI}_{t-1}} \times 100. \tag{1}$$

for a given month t. The result is given in Figure 1.

2.1.2. Google trends data: For the specific keywords used in searches, the data are obtained from Google Trends website [48]. We focused on a precise set of queries directly related to inflation rises. The baseline set of keywords was inspired by the Keywords Everywhere website [49]. We selected 105 terms that reflect Moroccans' reactions and capture trends similar to the inflation rate series during the studied period (see also Figure 2). We refined the choice of keywords by sorting them into categories, each category linked to searches for information related to inflation. These categories guide the selection of keywords in different languages, ensuring comparability. The usefulness of the study may vary depending on the environment in which judgments about intentions are made using search volume data. Search queries are often used to gather information about certain topics. When a topic is exclusively of interest to persons in specific settings, variations in search volume may imply an increased incidence of those terms in that context. To this end, the keywords are selected so that they show a substantial increase in search traffic that occurs at virtually the same time as the rate of inflation over the month.

The “Energy” category includes queries about electricity and oil commodities, while, the “Finance” category encompasses queries about foreign currency and trade transactions. The ”Food” category is broader and may be used by individuals not directly affected by inflation, though it is still relevant to the context of inflation in Morocco. Additionally, the ”Website” category captures general attention to inflation-related topics, whereas, the ”Other Keywords” category includes terms likely to be part of specific sentence queries, such as ”price of (a product)”.

2.2. Methodology

The present study begins with a Data Analysis phase. First, We uploaded Consumer Price Index data and conducted a comprehensive collection of data of Google Trends data. Then, these data were

Langage	Keywords
Arabic	بقر, الفلاحة, توصيل مجاني, البنزين, ضريبة, الزراعة, النفط, فواكه, برنت, التضخم, برميل النفط, الطاقة, المازوت, الاسعار, ارخص, السوق, الجفاف, اسعار المحروقات, استثمار, دواء, البترول, الخضر, غلاء, اسعار النفط, ثمن, عملة, صيدلية, فاتورة, سعر برميل, ثمن برميل النفط اليوم, الطاقة الشمسية, سعر البترول, المحروقات, لحم, النفط اليوم, النفط اليوم
French	alimentaire, agriculture, énergie , salaire, électricité, prix immobilier, marché, investissement, Taux d'intérêt, guichet, sécheresse, produit, gasoil, change, taux directeur, devise, pétrole, transport, oncf voyage, combustible, prix carburant maroc, gasoil prix maroc, gasoil maroc, vente, compte, taux de change, carburant
English	inflation, energy, product, house, insurance, dollar, tax, euro, salary, import, trade, fruit, usd to mad, bank, money, oil, balance, euro to mad, portfolio, dollar to mad, buy, sell, interest rate, rate, Amazon merch, Wise, Upwork, Fiverr, Etsy, Kiwi, Canva, Redbubble, Avito, Atacadao

Figure 2. Visualization of 105 searched keywords

subjected to a Unit Root Test to determine if they were stationary. In case of non-stationarity, we applied the differentiation. Once stationarity was achieved, a Granger Causality Test was performed to identify causality relationships, leading to the selection of relevant keywords for predicting the inflation index. Following this, we constructed a Google Trends index using only the most important keywords. Concurrently, the inflation data underwent the Box-Jenkins methodology to develop a Seasonal Autoregressive Integrated Moving Average (SARIMA) model. In the Linear Analysis phase, the PCA index and the SARIMA model were combined in a SARIMAX model to generate the final forecasting output (Figure 3).

2.2.1. Unit Root test: Data transformation is frequently necessary during the identification stage to ensure that the time series are stationary. Stationarity (or unit root) is required for developing an ARIMA model for forecasting process [50]. Before fitting an ARIMA model, differencing and power transformation of the data are performed to eliminate the trend and stabilize the variance when the observed time series exhibits heteroscedasticity and a trend. To check the stationarity of the (transformed) time series, the

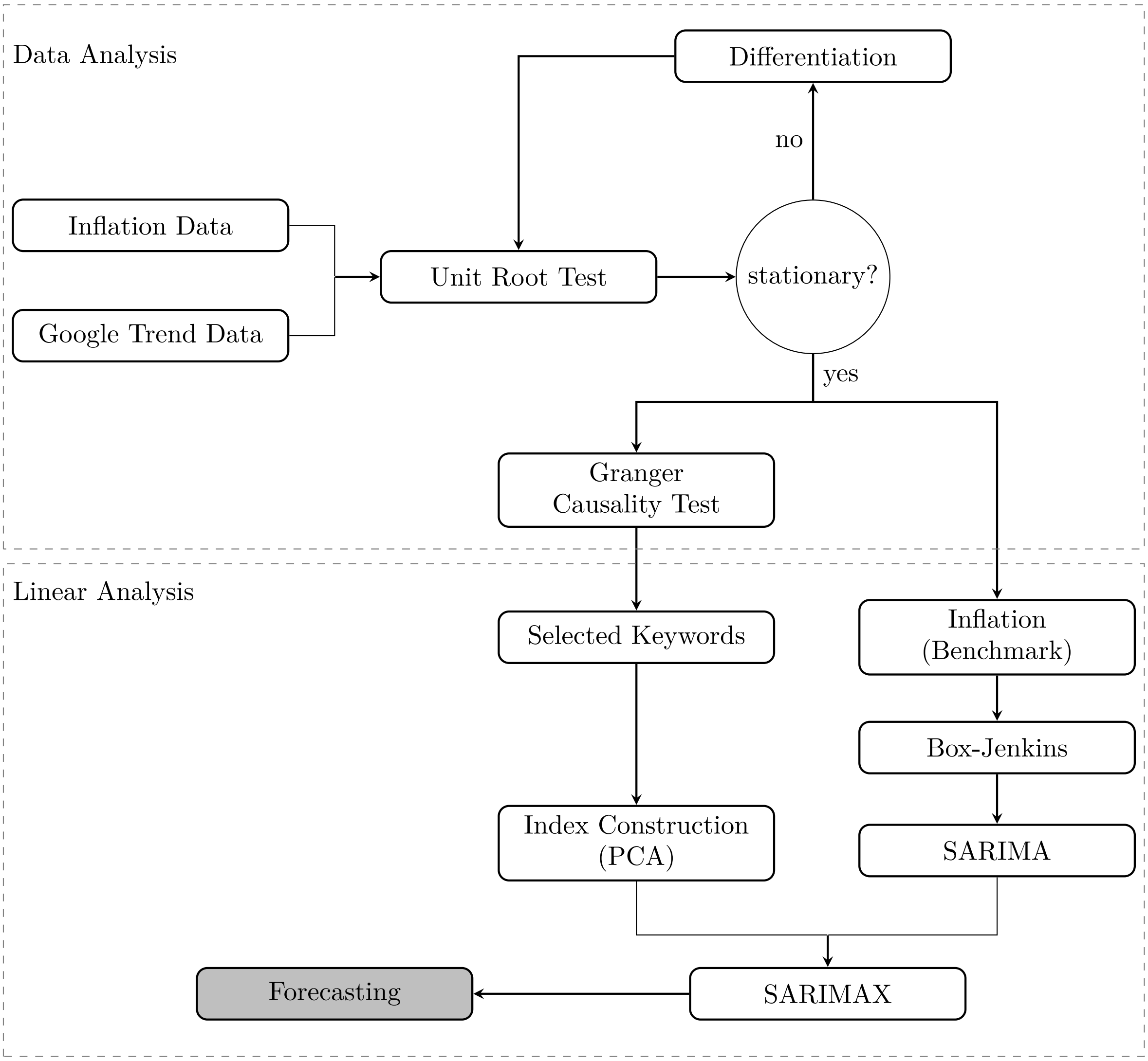

Figure 3. An overview of the study methodology

augmented Dickey-Fuller test (ADF) was applied. The ADF test equation is as follows:

$$\Delta y_t = \alpha_0 + \alpha_t + p y_t + \sum_{i=1}^{k} \gamma_j \Delta y_{t-1} + \epsilon_t. \tag{2}$$

As a result, a stationary time series have stable statistical features, including mean and autocorrelation structure, across time.

2.2.2. Granger Causality test: Typically, to test hypotheses regarding Granger causality, two linear autoregressive (AR) models for stationary time series y and z are assumed [51]. The first model predicts the current value of y based on p delayed values of both y and z. The second model predicts the current value of z using different coefficients. The regression errors in the following equations are denoted by ϵ

and the regression coefficients by a.

$$\begin{aligned} y(t) &= \sum_{j=1}^{p} a_{11,j} y(t-j) + \sum_{j=1}^{p} a_{12,j} z(t-j) + \epsilon_1(t) \\ z(t) &= \sum_{j=1}^{p} a_{21,j} y(t-j) + \sum_{j=1}^{p} a_{22,j} z(t-j) + \epsilon_2(t) \end{aligned} \tag{3}$$

In the first equation, if the variance of error ϵ_1 is significantly lower for the model with variable z (coefficients a_{12} different from zero) than the same model without variable z (coefficients a_{12} equal to zero), it indicates that z is the Granger cause of the variable y. Similar conclusion can be depicted from the second equation. In our case, we test the null hypothesis that the variable z (keywords) is Granger causing the variable y (inflation).

2.2.3. Principal Component Analysis (PCA): PCA is one of the most widely used methods for dimensionality reduction. In this study, we focus on PCA to initially assess the significance of the Google Trends index. Specifically, given the high-dimensional set of Google Trends queries (105 dimensions), PCA provides a standard method for identifying the optimal lower-dimensional subspace in terms of minimizing least-squares error. The first principal component captures the direction that minimizes the distance between the original data and its projection onto this component, thereby preserving as much variance as possible in a reduced form.

It is possible to appropriately modify the correlation matrix's representation of the dependency structure of search intensity across keywords in order to lower the dimensionality with the least amount of information loss [42]. With the use of PCA, orthogonal factors may be effectively extracted from a bigger information set such that the top factors account for the bulk of the variability in the information domain. However, since information loss is still a serious concern, include only one element is probably not the best option. Thus, the composite indicator may be computed as,

$$X_t = \frac{1}{\sum_i \nu_i} \sum_i (\nu_i k_t^i) \tag{4}$$

where k_t^i is the i^{th} extracted keyword at time t and ν_i is the proportional variability explained by i^{th} factor. The alternative model of the benchmark SARIMA model includes the first principal component as explanatory variable. This approach sets X_t as the principal component derived from the set of variables obtained from Google Trends data, where (N = 105).

2.2.4. Linear Modeling. The ARMA model combines the autoregressive (AR) and moving average (MA) processes. The current value of the time series for the AR model of order p, referred to as a AR(p) model, may be stated using the standard formula that follows,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \tag{5}$$

whereas MA(q) model presents the current and q prior values of random errors, namely,

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}. \tag{6}$$

Consequently, we can construct the generic formula for a coupled ARMA(p, q) process as follows

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}, \tag{7}$$

where, for each previous observed value, y_{t-i}, ϕ_i are coefficients associated with the value predicted or simulated, the coefficients θ_i are linked to past white noises, ε_t is a typical white noise process, with a variance of σ^2 and a mean of zero.

The ARMA model is commonly used to analyze stationary time series. Nevertheless, if the series are nonstationary, they are converted into a stationary time series through d^{th} difference process. In most cases, d is equal to $0, 1$, or a maximum of 2 [52]. Consequently, it is possible to get the ARIMA(p, d, q) in the following form:

$$w_l = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \cdots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{8}$$

where $w_t = \nabla^d y_r$.

The SARIMA model eliminates non-stationarity of the series by suitable seasonal differencing. The six components of SARIMA are Autoregressive (AR), Integration (I), Moving Average (MA), Seasonal Autoregressive (SAR), Seasonal Integration (SI), and Seasonal Moving Average (SMA). The matching parameters are, in that order, $p, d, q, P, D,$ and Q. For monthly time series $m = 12$, the standard SARIMA model is denoted as SARIMA $(p, d, q)(P, D, Q)[m]$, where m is the order of seasonal differencing (Table 1)

$$\phi_p(B) \cdot \Phi_P\left(B^m\right) \cdot \nabla^d \cdot \nabla_m^D \cdot Y_T = \theta_q(B) \cdot \Theta_Q\left(B^m\right) \cdot \epsilon_t \tag{9}$$

where $\nabla^d \cdot \nabla_m^D \cdot Y_I$ is the seasonally differenced series, B is the back-shift operator.

Table 1. Definition of SARIMA parametrs

Name	Order	Operator	Equation
Autoregressive (AR)	p	$\phi_p(B)$	$1 - \phi_1 B^1 - \phi_2 B^2 \ldots - \phi_p B^p$
Integration (I)	d	∇^d	$\left(1 - B^1\right)^d$
Moving Average (MA)	q	$\theta_q(B)$	$1 - \theta_1 B^1 - \theta_2 B^2 \ldots - \theta_q B^q$
Seasonal AR (SAR)	P	$\Phi_p\left(B^m\right)$	$1 - \Phi_1 B^{1 \cdot m} - \Phi_2 B^{2 \cdot m} \ldots - \Phi_p B^{P \cdot m}$
Seasonal Integration (SI)	D	∇_m^D	$\left(1 - B^m\right)^D$
Seasonal MA (SMA)	Q	$\Theta_Q\left(B^m\right)$	$1 - \Theta_1 B^{1 \cdot m} - \Theta_2 B^2 \cdot \ldots, \ldots - \Theta_Q B^{Q \cdot m}$
Back Shift	B	B	$B^n \cdot Y_t = Y_{t-n}$

Crucially, the model we present is designed to generate inflation projections by leveraging information from a comparable long-memory using principal component time series. While ARIMA models are a more traditional approach for analyzing and forecasting non-stationary time series data, our model offers a novel alternative. In fact, an exogenous main component variable is created by matching the monthly observed data with the inflation estimates derived from the Google search volume data. Consequently, the augmented ARIMAX (p, d, q) model can be expressed as follows:

$$y_t = \mu + \gamma X_t + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t. \tag{10}$$

Here, X represents the time series consisting of the constructed principal component index for the relative selected terms, and γ is the coresponding parameter. In principle, the parameters in Eq. (10) could be directly estimated by OLS. As will be seen in Sect. 3.2, it turns out that the models did not require all of the variables. The assessment will be based on the p values of every variable, determining if the addition of one or more model variables causes a substantial decline in the performance indices (AIC and RMSE, as established in Sec. 2.2.5).

2.2.5. Evaluation Criteria of Forecasting Performance: The Box-Jenkins approach [52] consists of three iterative steps: model identification, parameter estimation, and diagnostic check. Model identification is based on the basic principal that a time series produced by an ARIMA process ought to possess certain theoretical autocorrelation characteristics. Finding one or more viable models for a particular time series may frequently be accomplished by comparing the theoretical and actual autocorrelation patterns. Box and Jenkins suggested using the sample data's autocorrelation function (ACF) and partial autocorrelation function (PACF) as the fundamental instruments for determining the ARIMA model's order. Other validity-based order selection techniques can be considered, such as information-theoretic techniques like Akaike's information criterion (AIC),

$$\mathrm{AIC} = N \log \frac{\sum_{j=1}^{N} (y_j - \hat{y}_j)^2}{N} + 2K + \frac{2K(K+1)}{N-K-1} \tag{11}$$

where K is the number of parameters of the model, and the Root mean Square Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{N} (y_j - \hat{y}_j)^2}{N}}. \tag{12}$$

It is possible to assess the tentatively entertained model's quality of fit to the historical data using a number of diagnostic statistics and residuals plots. In case that the model proves insufficient, it is advisable to find a new, provisional model and proceed with the parameter estimates and model verification procedures once more. Alternative model(s) may be suggested with the use of diagnostic data. Usually, several iterations of this three-step model development procedure are conducted until a suitable model is ultimately chosen. Predictions can then be made using the final model that was chosen.

3. Results and Discussion

In this section, we highlight the key findings from our modeling method. As previously mentioned, the three primary phases of the ARMA model are, model identification, parameter estimation, and application. The identification step, which consists of two stages, is the most crucial of the three: In order to attain stationarity and normality, (1) the series are appropriately differencing if needed, and (2) the orders of the AR and MA elements of the ARMA model are determined.

3.1. Unit Root test

Actually, all data series are seasonally adjusted and differentiated to achieve stationarity. Subsequently, Augmented Dickey-Fuller (ADF) tests are used to check for stationarity (lag length selection depending on the BIC). We reject the unit root null hypothesis in favor of the stationary alternative. Table 2 provides inflation data along with a descriptive statistical illustration (identical ADF findings are produced for the selected keywords from Figure 2).

Table 2. Descriptive statistics for the inflation rate series from January 2018 to September 2023

Variable	Min	Max	Mean	Std. dev.	ADF test
Inflation	−0.6064	10.0848	2.7941	2.9422	−2.2002
dInflation	−1.8440	1.6498	0.0398	0.6746	−2.7715
ddInflation	−3.0189	1.7736	0.0034	0.8684	−5.8607***

Significance code. '***' : 0.001, '**' : 0.01, '*' : 0.05 , . : 0.1, ' ' : 1

We note also from Table 2, mean ($\simeq 0$) and standard deviation ($\simeq 1$) values that the stationary inflation tends to have a standard normal distribution.

Granger causality test between 18 GT terms and the monthly inflation was applied (Figure 4). Moreover, the corresponding search volume indices for the 18 distinct phrases are shown visually in

Keyword	Tanslation	F-test	P-value	Lags
prix immobilier	property prices	17.512	0.0000***	1
import	import	8.3671	0.0052**	1
rate	rate	8.2411	0.0055**	1
etsy	etsy	7.7362	0.0071**	1
redbubble	redbubble	5.8742	0.0182*	1
pétrole	oil	6.5848	0.0126*	1
atacado	atacadao	5.7158	0.0198*	1
البترول	oil	5.2656	0.0251*	1
المحروقات	Hydrocarbons	4.5371	0.0370*	1
market	market	4.8125	0.0319*	1
فاتورة	bill	4.147	0.0459*	1
سعر برميل النفط اليوم	oil price per barrel today	3.5863	0.0628.	1
upwork	upwork	3.4851	0.0665.	1
بقر	beef	3.4118	0.0694.	1
سعر البترول	oil price	3.4002	0.0699.	1
النفط	oil	3.1041	0.0829.	1
transport	transport	2.9811	0.0891.	1
فواكه	fruits	3.0016	0.0889.	1

Significance code. ‘***’ : 0.001, ‘**’ : 0.01, ‘*’ : 0.05 , . : 0.1, ‘ ’ : 1

Figure 4. Granger causality test between 18 GT terms and the monthly inflation series

the correlation matrix (heatmap) in Figure. 5. There is a significant degree of connection between different phrases, which is to be expected considering that Google will automatically suggest similar search terms to consumers when they type a single keyword. Searches for particular phrases are therefore likely to be connected with searches for other terms because many of these words are associated with economics-related issues from three distinct languages.

Interestingly, we created a sample Google Trends index using principal component analysis (PCA). As a matter of fact, PCA is the one that reduces the gap between the data and its projection. It will attempt to maintain as much variety as feasible while decreasing the dimensionality of the chosen 18 keywords. The original variables are converted into a new collection of uncorrelated variables, with the first few variables being arranged to preserve the majority of the variance found in the original dataset.

3.2. Model selection

Subsequent to the foregoing analysis, we use correlation functions of inflation data as a first method to determine an appropriate lag structure. In fact, The primary instruments used by Box and Jenkins to determine the ARIMA model's order were the sample data's autocorrelation function (ACF) and partial autocorrelation function (PACF). More specifically, the ACF has a cut-off at lag q if the sample data is an MA(q) process, and the PACF cut-off is at lag p if the sample data is an AR(p) model.

Figure 6 depicts the auto correlations at different lags of inflation changes. The series feature a high persistence as correlations change slowly between lags. It can be noticed that the autocorrelation

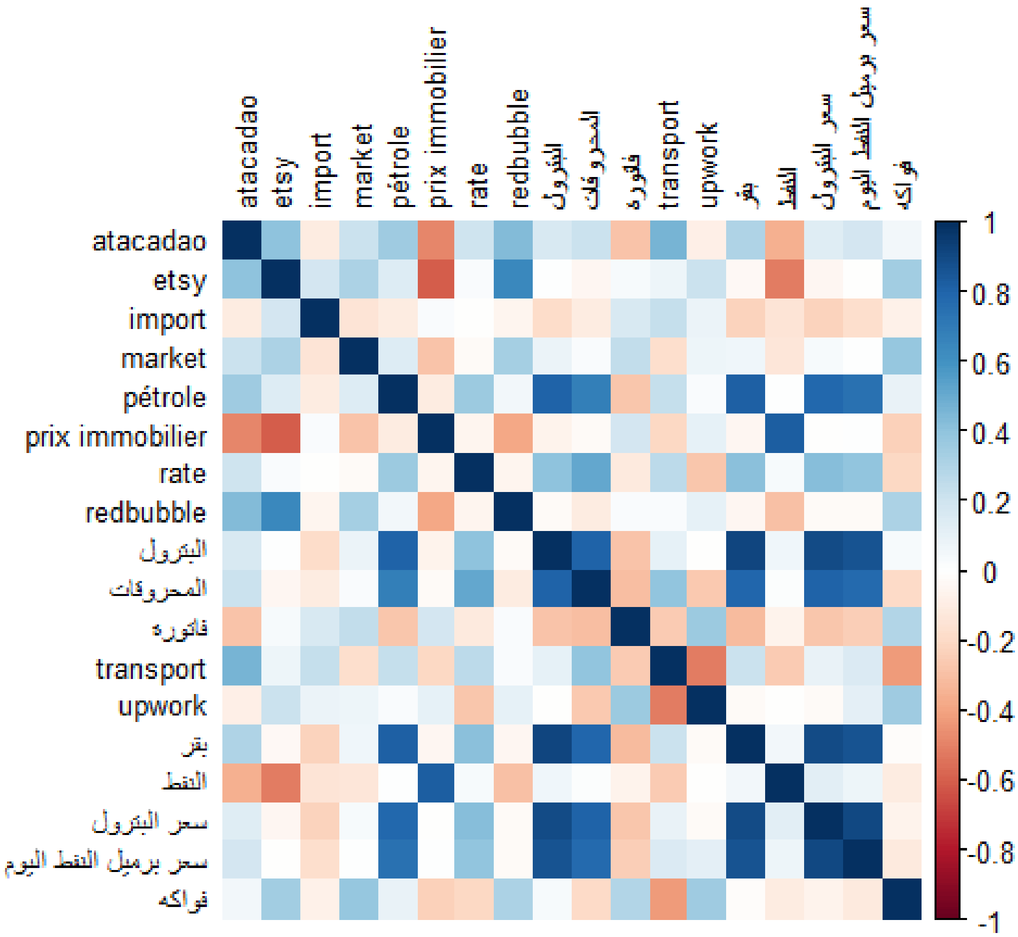

Figure 5. Visualization of the heat-map of the 18 selected keywords

coefficient is significantly different from 0 in lag 1 and it is basically in the confidence band when the lag order is higher than 1. Thus q can be equal to 0 or 1. The partial autocorrelation coefficient is significantly different from 0 when the lag order equals to 1 and 2. So, $p = 0$, $p = 1$, and $p = 2$ can be examined. The value of Q is deduced from the ACF values which presents a seasonality after a period of 1 year (12 months). Thereby, $Q = 0$ or $Q = 1$.

Observing the ACF and PACF graphs does not provide obvious values for p and q in mixed models. Other identification techniques, such the RMSE and AIC shown in Sec. 2.2.5, have been proposed based on information-theoretic methodologies. The best fitting model is chosen in this study using the RMSE and AIC criteria. After a suitable model has been selected, the model's parameters need to be estimated

The values of the various models are summarized in Table 3, which also supports the model selection that will serve as the foundation for our projections. In the context of autoregressive coefficient analysis, the SAR(1) is only significant for models including the MA(1) variable, whereas the AR(1) yields a significant and negative coefficient. The projected seasonal moving average SMA(1), on the other hand, is consistently negative and extremely significant. The models' stability and robustness are demonstrated by all of these values, whose statistics and coefficients do not change noticeably when more explanatory variables are included.

Table 3. Estimated coefficients from different models

Models	Cases	Estimated coefficients						AIC	RMSE
		Intercept	AR(1)	SAR(1)	MA(1)	MA(2)	SMA(1)		
SARIMA(1,0,0)(1,0,0)	zero mean	-	-0.4577^{***}	-0.49334^{***}	-	-	-	149.65	0.6881
	no zero mean	-0.0008	-0.4577^{***}	-0.4934^{***}	-	-	-	151.65	0.6881
SARIMA(0,0,1)(0,0,1)	**zero mean**	-	-	-	-0.8625^{***}	-	-0.4947^{***}	**133.45**	**0.6038**
	no zero mean	-0.0004	-	-	-0.8608^{***}	-	-0.4942^{***}	135.44	0.6039
ARIMA(0,0,1)	zero mean	-	-	-	-0.8495^{***}	-	-	144.23	0.6824
	no zero mean	-0.0036	-	-	-0.8477^{***}	-	-	0.6822	146.16
SARIMA(0,0,1)(1,0,1)	zero mean	-	-	-0.2446	-0.8593^{***}	-	-0.3000	134.72	0.5997
	no zero mean	0.0000	-	-0.2445	-0.8595^{***}	-	-0.3001	136.72	0.5997
SARIMA(0,0,1)(1,0,0)	zero mean	-	-	-0.4723^{***}	-0.8526^{***}	-	-	133.99	0.6082
	no zero mean	-0.000	-	-0.4717^{***}	-0.8511^{***}	-	-	135.98	0.6083
SARIMA(0,0,0)(0,0,1)	zero mean	-	-	-	-	-	-0.4348^{***}	164.08	0.7841
	no zero mean	0.0021	-	-	-	-	-0.4347^{***}	166.07	0.7841
SARIMA(1,0,1)(0,0,1)	zero mean	-	0.0929	-	-0.9059^{***}	-	-0.4896^{***}	135.11	0.6011
	no zero mean	0.0028	0.1436	-	-0.9999^{***}	-	-0.4883^{***}	136.86	0.5868
SARIMA(0,0,2)(0,0,1)	zero mean	-	-	-	-0.8128^{***}	-0.0786	-0.4899^{***}	135.12	0.6014
	no zero mean	0.0028	-	-	-0.8670^{***}	-0.1329	-0.4911^{***}	136.98	0.5868
SARIMA(1,0,0)(0,0,1)	zero mean	-	-0.4545^{***}	-	-	-	-0.4703^{***}	150.65	0.6951
	no zero mean	-0.0012	-0.4546^{***}	-	-		-0.4704^{***}	152.65	0.5354
SARIMA(1,0,2)(0,0,1)	zero mean	-	-0.8370^{***}	-	0.1204	-0.8795^{***}	-0.4878^{***}	134.21	0.5775
	no zero mean	-0.0002	-0.8371^{***}	-	0.1215	-0.8784^{***}	-0.4874^{***}	136.21	0.5775

Std. errors are displayed in parenthesis. Significance code. '***' : 0.001, '**' : 0.01, '*' : 0.05 , ' ' : 1

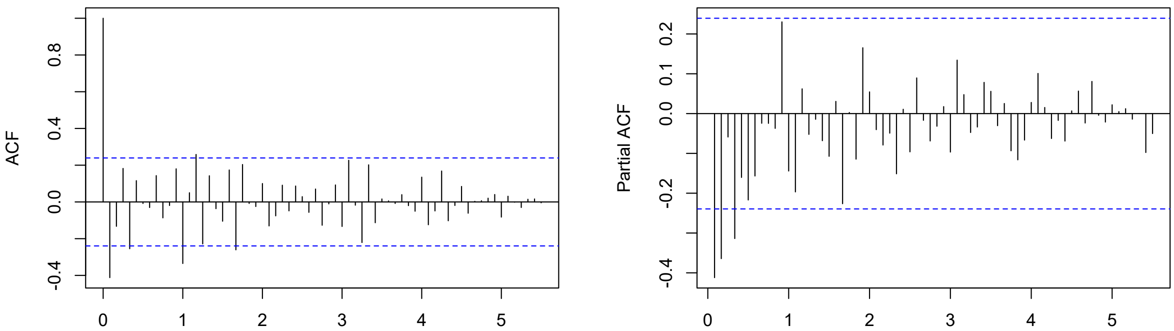

Figure 6. The figure presents the partial auto correlation function of the current change in inflation rate with different lags. The left graph depicts the auto correlation for Morocco while the right graph presents partial autocorrelation. The dotted horizontal lines represent 95% confidence intervals

Table 3 indicates that the SARIMA$(0, 0, 1)(0, 0, 1)$ model was chosen due to its significant coefficients and appropriate amount of adjustment. As such, the ideal model prioritizes simplicity and reduces AIC and RMSE evaluation metrics.

3.3. Forecasting results

We conduct the predictions after identifying the inflation model that best fits our situation. We use the Principal Component Index to improve trend prediction and forecasting. Table 4 and Figure 7 show the inflation projections produced from our model for the period January 2018 to September 2023. The estimated parameters for SARIMA and SARIMAX are shown in Table 4.

In comparison to PCA-based and benchmark models, the feature selection-based models have much lower RMSE , AIC and Log-likelihood criteria, and this difference diminishes with each new GT query that explains the data. This was anticipated due to the way the feature selection method (Granger Causality) was created.

Table 4. Results of SARIMAX model compared to the selected model in Table 3

Parameter	Symbol	SARIMA	SARIMAX
MA(1)	θ	-0.8625^{***}	-0.8685^{***}
SMA(1)	Θ	-0.4947^{***}	-0.6537^{***}
X	γ	-	0.1231^{***}
AIC		133.45	125.14
RMSE		0.6038	0.5451
Log likelihood		-63.72	-58.57

Significance code. ‘***’ : 0.001, ‘**’ : 0.01, ‘*’ : 0.05 , . : 0.1, ‘ ’ : 1

Figure 7 makes it abundantly evident that the selected model is capable of modeling and projecting Morocco’s inflation in the future. The decision-making process was made easier by the forecasts that were produced following our modeling approach. Indeed, using specific keywords and public preferences, the algorithm allowed us to anticipate inflation and provide precise forecasts. Upon obtaining an inflation

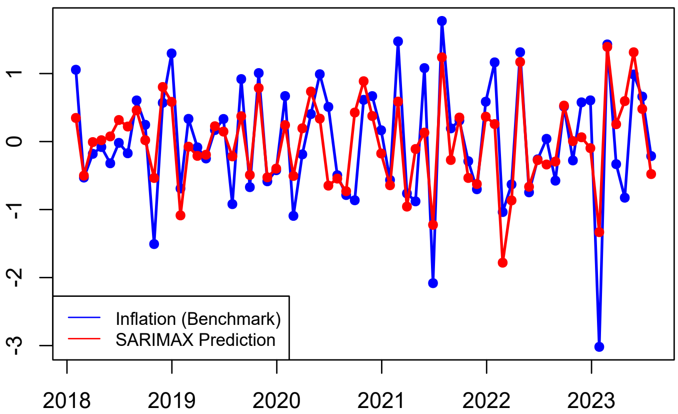

Figure 7. Forecasts from an SARIMAX model fitted to the monthly inflation data from January 2018 to September 2023 for Morocco

prediction, making informed decisions and averting anticipated issues will become considerably simpler and more evident.

Overall, the model utilizing the Google Search Index had the most captivating and outstanding results. There are two key benefits to the PCA index. The first benefit is a decrease in multicolinearity, as Google Keywords in Morocco have a high degree of correlation due to the country's diversity of languages. Flexibility is the second benefit, since the data may be readily transformed into a single time series for conventional economic research needs, including predicting a long-term connection at a monthly frequency.

3.4. Robustness Analysis & Final remarks

To validate the forecasting performance of the feature selection-based model (SARIMAX) demonstrated in the previous section, we carry out a robustness analysis specifically for this model. This involves examining the impact of different train-test data configurations that could influence forecasting results, including the specification sample and the forecasting period.

With a few exceptions, the results shown in Figure 8 are similarly significant: the use of Google Trends (GT) queries combined with the proposed feature selection-based model improves the forecasting accuracy on the test data. The simulated trajectories (represented in gray) encompass most of inflation data collected over the selected period, ensuring a comprehensive analysis of inflation trends and providing a robust foundation for understanding the model's underlying dynamics. By integrating all relevant data, we enhance the accuracy and reliability of our simulations. This, in turn, highlights the performance of SARIMAX model and the importance of GT index in addressing future inflation expectations.

Although Google Trends has been a valuable tool for economic aggregates, it has various limitations that require careful consideration. Addressing these limitations is essential for accurate interpretation. Actually, Google Trends captures only the interests and behaviors of individuals with internet access and utilizing the search engine run by Google. Thus, It may not accurately represent the interests and behaviors of regions with restricted internet access and a low presence of Google users. Additionally, a recent review by [53] has indicated that Google Trends may have restricted validity in regions with low internet access or free expression. Therefore, it is advisable to use traditional techniques, like surveys, to better capture public preferences and issues, particularly when the investigation intends to understand the interests of the entire population, encompassing non-internet users. Likewise, other search engines, like Baidu, Bing, Yahoo, and Yandex, could be considered in regions where Google Search is less popular. Moreover, Google Trends can be combined with data from other widely used online platforms, like

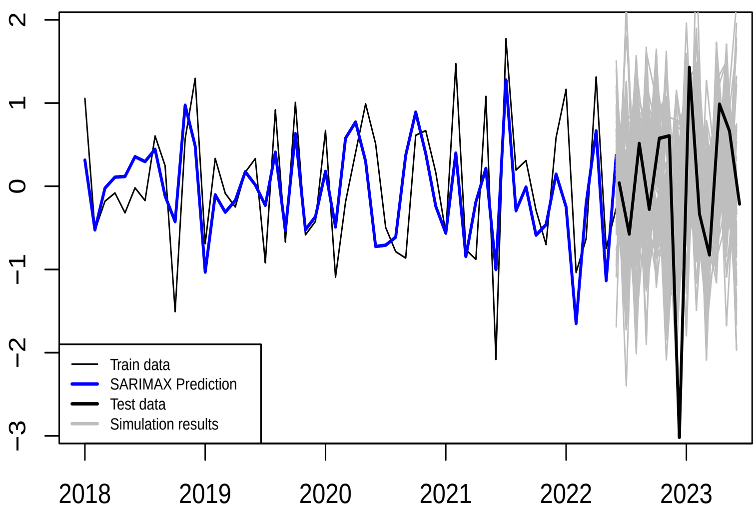

Figure 8. Forecasts from an SARIMAX model fitted to the monthly inflation data from January 2018 to September 2023 for Morocco using train-test data

Wikipedia, to gain deeper insight into online public interests and behaviors. Accordingly, it is noteworthy that Google Trends reports relative values instead of absolute volume of searches and requests. Thereby, the exact number of queries for a specific topic remains unknown. Besides, Google samples its data to create a Google Trends dataset comprising the entire Google searches for a specific topic. Nevertheless, the sampled queries remain undisclosed. Consequently, the data may show slight variations depending on when it is retrieved [54]. Additionally, Google leverages artificial intelligence to aggregate search queries. Nevertheless, the terms and phrases included under topic keywords in Google Trends are not specified because of restricted transparency of Google Trends' algorithm [55]. Consequently, to enhance replicability and effectiveness, prior research has suggested several strategies, like applying statistical methods [54] to gather average search volumes over multiple days, or as suggested before, observing key methodological settings like query category, the exact keywords, use of quotation marks, the rationale for choosing keyword, time intervale, date, and location of retrieval. Besides, shifts in search behavior related to particular topics may stem from various factors, such as greater media interest [56]. Consequently, researchers should clearly outline their assumptions regarding search habits and attentively interpret Google Trends data, while taking into account its limitations.

4. Conclusion & Perspectives

Our work's results show that predicting accuracy may be successfully increased by adding more data, which can be gathered in the form of time series from search queries. Essentially, by using Google search data, we are able to build a SARIMAX model that predicts changes in inflation rates. In this setting, very basic linear models perform remarkably well, and our suggested strategies can produce large external factors over the analyzed time. The suggested approach has a number of benefits. First, there is a greater likelihood of locating the ideal lag orders for a near-perfect forecast when using Box-Jenkins and selection criteria approaches. Second, although the Granger Causality's fundamental idea is straightforward, it may offer profound insight into the features of Google Trends time series. Third, the dimensional reduction approach (PCA) is useful for analyzing a large external element in the investigation. The research also sheds some insight on the discussion over the use of explanatory variables in the literature. Lastly, there is no need to make difficult choices about the explicit form of the suggested model. In light of this, using a hybrid forecasting model with Google Trends data may produce forecasts that are more accurate and stable. It may also prove useful in research pertaining to economic time series forecasting for a variety of issues including macroeconomic aggregates.

Theoretically, adding more terms may improve the performance of the different selected linear models, as some of the dynamics inside real movements may not be entirely captured by our existing terms. Furthermore, behavioral data from several sources has been maintained reasonably separate in a large portion of the current work in this field. Combining Twitter sentiment analysis data with the already-existing Google data might be a fascinating research to develop a model that captures various dynamics inside the behavioral inefficiencies. Additional developments for this research involve adjusting the data's frequency, for example, by using a mixed-frequency nowcasting strategy [57, 58]. In reality, Google Trends data is available on a daily, weekly, monthly, and even hourly basis. We have only examined monthly data thus far. According to our first investigation, relatively few keywords that Granger cause inflation rate are produced when using Google Trend data collected once a month. Investigating this relationship's dissociation at a finer data aggregation level might be worthwhile.

References

1. R. Ibarra, *Do disaggregated CPI data improve the accuracy of inflation forecasts?*, Economic Modelling, vol. 29, no. 4, pp. 1305–1313, 2012.
2. J. Xiao, M. Wang, L. Tian, and Z. Zhen, *The measurement of China's consumer market development based on CPI data*, *Physica A: Statistical Mechanics and its Applications*, vol. 490, pp. 664–680, 2018. Elsevier. doi: https://doi.org/10.1016/j.physa.2017.08.135.
3. J. Faust and J. H. Wright, *Forecasting Inflation*, in *Handbook of Economic Forecasting*, vol. 2, Elsevier, pp. 2–56, Jan. 2013.
4. S. El Khamlichi, A. Maurady, and A. Sedqui, *Comparative study of COVID-19 situation between lower-middle-income countries in the eastern Mediterranean region*, Journal of Oral Biology and Craniofacial Research, vol. 12, no. 1, pp. 165–176, 2022.
5. M. Woodford, *Inflation targeting and financial stability*, National Bureau of Economic Research, Tech. Rep. No. w17967, 2012.
6. S. Gilchrist, R. Schoenle, J. Sim, and E. Zakrajšek, *Inflation dynamics during the financial crisis*, American Economic Review, vol. 107, no. 3, pp. 785–823, 2017. doi: https://doi.org/10.1257/aer.20150248.
7. O. Armantier, G. Koşar, R. Pomerantz, D. Skandalis, K. Smith, G. Topa, and W. Van der Klaauw, *How economic crises affect inflation beliefs: Evidence from the Covid-19 pandemic*, Journal of Economic Behavior & Organization, vol. 189, pp. 443–469, 2021. doi: https://doi.org/10.1016/j.jebo.2021.04.036.
8. O. Barkan, J. Benchimol, I. Caspi, E. Cohen, A. Hammer, and N. Koenigstein, *Forecasting CPI inflation components with hierarchical recurrent neural networks*, *International Journal of Forecasting*, vol. 39, no. 3, pp. 1145–1162, 2023. Available at: https://doi.org/10.1016/j.ijforecast.2022.04.009.
9. M. Friedman, *The lag in effect of monetary policy*, Journal of Political Economy, vol. 69, no. 5, pp. 447–466, 1961, The University of Chicago Press. doi: https://doi.org/10.1086/258537.
10. T. Cogley, S. Morozov, and T. J. Sargent, *Bayesian fan charts for UK inflation: Forecasting and sources of uncertainty in an evolving monetary system*, Journal of Economic Dynamics and Control, vol. 29, no. 11, pp. 1893–1925, 2005. doi: https://doi.org/10.1016/j.jedc.2005.06.005.
11. S. Altug and C. Çakmaklı, *Forecasting inflation using survey expectations and target inflation: Evidence for Brazil and Turkey*, International Journal of Forecasting, vol. 32, no. 1, pp. 138–153, 2016. doi: https://doi.org/10.1016/j.ijforecast.2015.03.010.
12. D. Ida, *Sectoral inflation persistence and optimal monetary policy*, Journal of Macroeconomics, vol. 65, p. 103215, 2020. Elsevier. doi: https://doi.org/10.1016/j.jmacro.2020.103215.
13. Z. Mandalinci, *Forecasting inflation in emerging markets: An evaluation of alternative models*, International Journal of Forecasting, vol. 33, no. 4, pp. 1082–1104, 2017. doi: https://doi.org/10.1016/j.ijforecast.2017.06.005.
14. I. C. Baciu, *Stochastic models for forecasting inflation rate. Empirical evidence from Romania*, *Procedia Economics and Finance*, vol. 20, pp. 44–52, 2015. Elsevier. doi: https://doi.org/10.1016/S2212-5671(15)00045-3.
15. M. Iqbal and A. Naveed, *Forecasting inflation: Autoregressive integrated moving average model*, *European Scientific Journal*, vol. 12, no. 1, pp. 83–92, 2016. Available at: 10.19044/esj.2016.v12n1p83.
16. R. Fahrudin and I. D. Sumitra, *Forecasting Inflation Using Seasonal Autoregressive Integrated Moving Average Method for Estimates Decent Living Costs*, in *IOP Conference Series: Materials Science and Engineering*, vol. 662, no. 2, pp. 022062, 2019. IOP Publishing. doi: 10.1088/1757-899X/662/2/022062.
17. C. P. Barros and L. A. Gil-Alana, *Inflation forecasting in Angola: a fractional approach*, *African Development Review*, vol. 25, no. 1, pp. 91–104, 2013. Wiley Online Library. doi: https://doi.org/10.1111/j.1467-8268.2013.12016.x.

18. K. Carstensen and J. Hawellek, *Forecasting inflation from the term structure*, Review of World Economics, vol. 139, pp. 306–323, 2003.
19. J. H. Stock and M. W. Watson, *Forecasting output and inflation: The role of asset prices*, *Journal of Economic Literature*, vol. 41, no. 3, pp. 788–829, 2003. American Economic Association. doi: 10.1257/002205103322436197.
20. S. Jere and M. Siyanga, *Forecasting inflation rate of Zambia using Holt's exponential smoothing*, *Open Journal of Statistics*, vol. 6, no. 2, pp. 363–372, 2016. Available at: https://doi.org/10.4236/ojs.2016.62031.
21. N. Hauzenberger, F. Huber, and K. Klieber, *Real-time inflation forecasting using non-linear dimension reduction techniques*, *International Journal of Forecasting*, vol. 39, no. 2, pp. 901–921, 2023. Available at: https://doi.org/10.1016/j.ijforecast.2022.03.002.
22. T. Nyoni, *Modeling and forecasting inflation in Zimbabwe: A generalized autoregressive conditionally heteroskedastic (GARCH) approach*, 2018. Available: https://mpra.ub.uni-muenchen.de/88132/.
23. M. Bikourne, K. Akdim, and A. Ez-Zetouni, *L'effet des fluctuations des prix du pétrole sur la dynamique de l'inflation au Maroc: Une approche de modélisation stochastique*, *Les Cahiers du Plan*, no. 56, pp. 40–48, 2023. Available at: https://doi.org/10.34874/PRSM.cahiers-du-plan-i56.42537
24. M. Bikourne, K. Akdim, A. Khellaf, and A. Ez-Zetouni, *Investigating stochastic volatility and jumps in inflation dynamics: an empirical evidence with oil price effect*, *The European Physical Journal Plus*, vol. 138, no. 12, pp. 1142, 2023. Springer. Available at: https://doi.org/10.1140/epjp/s13360-023-04778-5
25. D. Delle Monache and I. Petrella, *Adaptive models and heavy tails with an application to inflation forecasting*, *International Journal of Forecasting*, vol. 33, no. 2, pp. 482–501, 2017. Available at: https://doi.org/10.1016/j.ijforecast.2016.11.007.
26. J. H. Stock and M. W. Watson, *Implications of dynamic factor models for VAR analysis*, National Bureau of Economic Research Cambridge, Mass., USA, 2005. Available at: 10.3386/w11467.
27. M. Forni, M. Hallin, M. Lippi, and L. Reichlin, *The generalized dynamic factor model: one-sided estimation and forecasting*, *Journal of the American Statistical Association*, vol. 100, no. 471, pp. 830–840, 2005. Available at: https://doi.org/10.1198/016214504000002050.
28. E. S. Silva, H. Hassani, and J. Otero, *Forecasting inflation under varying frequencies*, *Electronic Journal of Applied Statistical Analysis*, vol. 11, no. 1, pp. 307–339, 2018.
29. V. Ülke, A. Sahin, and A. Subasi, *A comparison of time series and machine learning models for inflation forecasting: empirical evidence from the USA*, *Neural Computing and Applications*, vol. 30, pp. 1519–1527, 2018. Available at: 10.1007/s00521-016-2766-x.
30. E. Nakamura, *Inflation forecasting using a neural network*, Economics Letters, vol. 86, no. 3, pp. 373–378, 2005. doi: https://doi.org/10.1016/j.econlet.2004.09.003.
31. A. Haider and M. N. Hanif, *Inflation forecasting in Pakistan using artificial neural networks*, *Pakistan Economic and Social Review*, pp. 123–138, 2009. https://www.jstor.org/stable/25825345
32. N.-T. Nguyen and T.-T. Tran, *Mathematical development and evaluation of forecasting models for accuracy of inflation in developing countries: a case of Vietnam*, *Discrete Dynamics in Nature and Society*, vol. 2015, no. 1, p. 858157, 2015. Available at: https://doi.org/10.1155/2015/858157.
33. X. Li, W. Shang, S. Wang, and J. Ma, *A MIDAS modelling framework for Chinese inflation index forecast incorporating Google search data*, *Electronic Commerce Research and Applications*, vol. 14, no. 2, pp. 112–125, 2015. Available at: https://doi.org/10.1016/j.elerap.2015.01.001.
34. I. Bouayad, J. Zahir, and A. Ez-zetouni, *Nowcasting and Forecasting Morocco GDP growth using Google Trends data*, *IFAC-PapersOnLine*, vol. 55, no. 10, pp. 3280–3285, 2022. Available at: https://doi.org/10.1016/j.ifacol.2022.10.129.
35. L. Monteforte and G. Moretti, *Real-Time Forecasts of Inflation: The Role of Financial Variables*, *Journal of Forecasting*, vol. 32, no. 1, pp. 51–61, 2013. Available at: 10.1002/for.1250.
36. M. Modugno, *Now-casting inflation using high frequency data*, *International Journal of Forecasting*, vol. 29, no. 4, pp. 664–675, Oct. 2013. Available at: 10.1016/j.ijforecast.2012.12.003.
37. E. S. Knotek II and S. Zaman, *Nowcasting U.S. Headline and Core Inflation*, *Journal of Money, Credit and Banking*, vol. 49, no. 5, pp. 931–968, 2017. Available at: 10.1111/jmcb.12401.
38. World Bank, *World Bank Data*, 2024. https://data.worldbank.org/.
39. H. A. Carneiro and E. Mylonakis, *Google trends: a web-based tool for real-time surveillance of disease outbreaks*, Clinical Infectious Diseases, vol. 49, no. 10, pp. 1557–1564, 2009, The University of Chicago Press.
40. A. Seifter, A. Schwarzwalder, K. Geis, and J. Aucott, *The utility of Google Trends for epidemiological research: Lyme disease as an example*, Geospatial Health, vol. 4, no. 2, pp. 135–137, 2010. doi: https://doi.org/10.4081/gh.2010.195.
41. M. Y. Huang, R. R. Rojas, and P. D. Convery, *Forecasting stock market movements using Google Trend searches*, Empirical Economics, vol. 59, pp. 2821–2839, 2020. Springer. doi: https://doi.org/10.1007/s00181-019-01725-1.
42. R. Mulero and A. García-Hiernaux, *Forecasting Spanish unemployment with Google Trends and dimension reduction techniques*, *SERIEs*, vol. 12, no. 3, pp. 329–349, 2021. Available at: 10.1007/s13209-021-00231-x.
43. E. Bantis, M. P. Clements, and A. Urquhart, *Forecasting GDP growth rates in the United States and Brazil using Google Trends*, International Journal of Forecasting, vol. 39, no. 4, pp. 1909–1924, 2023. Elsevier. doi: https://doi.org/10.1016/j.ijforecast.2022.10.003.
44. P. Molnár and M. Bašta, *Google searches and Gasoline prices*, in *Proceedings of the 2017 IEEE Eindhoven Energy Conference (EEM)*, pp. 1–5, Jun. 2017. Available at: 10.1109/EEM.2017.7981978.
45. J. Bleher and T. Dimpfl, *Knitting Multi-Annual High-Frequency Google Trends to Predict Inflation and Consumption*, Econometrics and Statistics, vol. 24, pp. 1–26, 2022. Elsevier. doi: https://doi.org/10.1016/j.ecosta.2021.10.006.

46. G. Guzman, *Internet search behavior as an economic forecasting tool: The case of inflation expectations*, *Journal of Economic and Social Measurement*, vol. 36, no. 3, pp. 119–167, 2011. Available at: 10.3233/JEM-2011-0342.
47. International Monetary Fund, *IMF Data*, 2024. https://data.imf.org/.
48. Google Trends, *Google Trends*, 2024. https://trends.google.com/trends.
49. Keywords Everywhere, *Keywords Everywhere*, 2024. https://keywordseverywhere.com/.
50. A. Fahim, Q. Tan, U. A. Bhatti, M. M. Nizamani, and S. A. Nawaz, *The nexus between higher education and economic growth in Morocco: an empirical investigation using VaR model and VECM*, *Multimedia Tools and Applications*, vol. 82, no. 4, pp. 5709–5723, 2023.
51. M. Rosoł, M. Młyńczak, and G. Cybulski, *Granger causality test with nonlinear neural-network-based methods: Python package and simulation study*, *Computer Methods and Programs in Biomedicine*, vol. 216, p. 106669, 2022.
52. G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden Day, 1976.
53. A. Mavragani and G. Ochoa, *Google Trends in infodemiology and infoveillance: methodology framework*, *JMIR Public Health and Surveillance*, vol. 5, no. 2, pp. e13439, 2019. JMIR Publications Inc., Toronto, Canada.
54. V. Z. Eichenauer, R. Indergand, I. Z. Martínez, and C. Sax, *Obtaining consistent time series from Google Trends*, *Economic Inquiry*, vol. 60, no. 2, pp. 694–705, 2022. Wiley Online Library.
55. S. V. Nuti, B. Wayda, I. Ranasinghe, S. Wang, R. P. Dreyer, S. I. Chen, and K. Murugiah, *The use of Google Trends in health care research: a systematic review*, *PLOS ONE*, vol. 9, no. 10, pp. e109583, 2014. Public Library of Science, San Francisco, USA.
56. R. L. Bach and A. Wenz, *Studying health-related internet and mobile device use using web logs and smartphone records*, *PLOS ONE*, vol. 15, no. 6, pp. e0234663, 2020. Public Library of Science, San Francisco, CA, USA.
57. T. Dimpfl and T. Langen, *How unemployment affects bond prices: a mixed frequency Google nowcasting approach*, Computational Economics, vol. 54, pp. 551–573, 2019. Springer. doi: https://doi.org/10.1007/s10614-018-9840-7.
58. H. Wei, Y. Guo, Z. Yu, and H. Cheng, *The impact of events on metal futures based on the perspective of Google Trends*, Resources Policy, vol. 74, pp. 102286, 2021. Elsevier. doi: https://doi.org/10.1016/j.resourpol.2021.102286.