MAKÜ | Journal of Mehmet Akif Ersoy University
Economics and Administrative Sciences Faculty

e-ISSN: 2149-1658
Volume: 12 / Issue: 1
March, 2025
pp.: 107-123

National Unemployment Rate Forecast with Google Trends

Savaş GAYAKER[1], Hasan TÜRE[2]

Abstract

The significant economic recession and ongoing COVID-19 pandemic are impacting various sectors. The decrease in employment, one of the main consequences of this economic stagnation, is felt intensely in Türkiye. The concern that today's unemployment problem will be experienced more intensely in the future brings to the fore studies on unemployment forecasting. To date, unemployment forecasting studies have received extensive coverage in the literature. This study aims to make more successful forecasts of unemployment data by using Google Trends (GT), which is frequently used in different fields today. Four GT-based variables were incorporated into traditional forecasting methods, including ARIMA, ARIMAX, and VAR models. The VAR GT3 model, which integrates GT data with annual inflation, provided the best forecasting performance among all tested models. The findings indicate that models incorporating GT data derived from various keywords yield more successful results than traditional models.

Keywords: Unemployment, Google Trends, VAR, ARIMA, ARIMAX.

1. Asst. Prof. Dr.,Ankara Hacı Bayram Veli University, savas.gayaker@hbv.edu.tr, https://orcid.org/0000-0002-7186-1532

2. Assoc. Prof. Dr., Ankara Hacı Bayram Veli University, hasan.ture@hbv.edu.tr, https://orcid.org/0000-0002-1975-9063

https://doi.org/10.30798/makuiibf.1446639

Article Type	Application Date	Admission Date
Research Article	March 3, 2024	March 25, 2025

To cite this article: Gayaker, S., and Türe, H. (2025). National unemployment rate forecast with Google Trends. *Journal of Mehmet Akif Ersoy University Economics and Administrative Sciences Faculty*, *12*(1), 107-123. https://doi.org/10.30798/makuiibf.1446639

1. INTRODUCTION

In recent years, the lack of employment opportunities in developing countries has been one of the biggest problems. Unemployment has become more deeply felt due to the global economic slowdown and the ongoing Covid-19 pandemic. However, perceived unemployment may vary according to the socioeconomic conditions of each country. At this point, dissatisfaction with past performance and concern about the future increase the interest of researchers and policymakers in studies on unemployment forecasts.

It is recognized that economic crises and structural changes mainly cause unemployment. In addition to these factors, demographic factors such as education, population growth, and migration also play a role as determinants of unemployment. In Türkiye, a significant migration issue that has been significantly impacted due to the conflict in Syria and Afghanistan, asylum seekers have negatively impacted the unemployment problem. Consequently, accurate unemployment prediction plays a pivotal role in shaping effective policy responses.

Predictive models can benefit by incorporating indicators reflecting individuals' job-search behavior or attitudes, potentially boosting forecast accuracy. For many years, variables such as individual behavior and attitudes have been measured through surveys, which are primary data methods. However, like many macroeconomic indicators, there may be delays in the publication of the unemployment rate indicators. Moreover, considering the reluctance of time-conscious individuals to fill out questionnaires, electronic data is expected to address this gap significantly.

Using web-based data derived from keyword searches has compelled researchers to contemplate maximizing its efficiency. To support this concept, there is a demand for up-to-date data to facilitate real-time forecasts of the unemployment rate (Fondeur & Karame, 2013). Today, the widespread use of the Internet and the proliferation of non-traditional data environments such as smartphones, smart sensors, digital media, Google, and web-based data sources have made it easier to obtain data on most of the everyday activities of companies and individuals. The electronic footprints left by individuals due to their Internet searches provide valuable information to researchers (Jun & Park, 2016). The digital data obtained through these unconscious footprints are inherently unbiased (Abraham et al., 2018; Ayyoubzadeh et al., 2020; Santillana et al., 2015), accurate (Han et al., 2012; Park et al., 2017; Santillana et al., 2015; Wilcoxson et al., 2020), and beneficial (Wu & Brynjolfsson, 2015; Vosen & Schmidt, 2011). They also offer more extensive, frequent, and current data than the more commonly employed surveys, typically subject to publication delays (Mulero & Garcia-Hiernaux, 2023). Roughly speaking, an unemployed person's unconscious disclosure of this situation by entering job search sites provides an indicator that can be used in unemployment forecasting. In short, the information collected from the Internet is crucial in providing information about the behavior of individuals on the Internet.

Google, one of the most significant search engines, provides users with such search statistics free of charge through its Google Trends (GT) tool. Google handles over 92 percent of all online searches worldwide (Lee, 2015). These statistics can be accessed by specifying the keywords related to the topic of interest and the period in which the words were searched. GT is one of the most widely used tools in applied economics literature among non-traditional data sources. (Cebrián & Domenech, 2023; Suhoy, 2009).

While online tools like Google offer a wealth of data, how can we put this to practical use? By tapping into the data from GT, we aim to see if it can help predict unemployment trends. This paper explores using Google Trends data to improve unemployment forecasting in Türkiye. Türkiye is a fascinating case to analyze because of its sharp rise in unemployment due to inflation and migration. Additionally, the occurrence of illegal migration impacts the accuracy of unemployment statistics. To provide a comprehensive approach to Turkish unemployment, we use simple exponential smoothing (SES), autoregressive moving average (ARMA), ARMA with explanatory variables (ARMAX), and vector autoregressive (VAR) models and also introduce Google searches for *job offers*, *jobs* as explanatory variables.

The paper is organized as follows: Section 2 provides a comprehensive literature review on utilizing GT as predictors, specifically focusing on its applications in predicting unemployment. Section 3 delves into the details of the data used in our analysis, particularly emphasizing the generation process of GT queries. The model and methodology used in our study are also presented in Section 3. Empirical Results are presented in Section 4. In Section 5, we compare the forecasting results of the proposed models across various combinations of GT data. Finally, the paper concludes with a discussion and presents some conclusive remarks.

2. LITERATURE REVIEW

GT data has attracted significant attention from researchers recently due to its ability to categorize Big Data sources according to time intervals, geographical locations, topics, and search terms. As more and more daily activities take place online, people inadvertently reveal their inclinations on these matters through their Internet searches (Belej, 2022; Blazquez & Domenech, 2018; Einav & Levin, 2014; Knipe et al., 2021; Rotter et al., 2021; Sherman-Morris et al., 2011; Yeh et al., 2018). Over the past two decades or so, studies in various fields have sought to demonstrate that models utilizing Internet-based search data can enhance prediction accuracy. In these studies, researchers aim to leverage unconventional data sources to forecast social, economic, or psychological behaviors and trends. Hassani and Silva (2015) explored the utilization of these data types across various fields.

In many forecasting studies, Internet-based data are included in the models as explanatory variables to improve forecasting performance. Mulero and Garcia-Hiernaux (2023) emphasize that

Google Trends data enable nowcasting models to provide more accurate forecasts than traditional indicators. Cebrián and Domenech (2023) investigate whether GT data can be reliable.

Internet-based data are now also integrated into many unemployment forecasting studies (Adu et al., 2023; Anvik & Gjelstad, 2010; Choi & Varian, 2012; D'Amuri & Marcucci, 2010; McLaren & Shanbhogue, 2011; Simionescu, 2020; Simionescu & Cifuentes-Faura, 2022a; Simionescu & Zimmermann, 2017). Unemployment is an important indicator for governments and researchers, especially considering the challenging situation in the Turkish labor market. This indicator is highly affected by migration from Syria and other countries in the region, the ongoing effects of the COVID-19 pandemic, and the economic slowdown.

The existing literature reveals that the number of studies focusing on Türkiye needs to be increased. Chadwick and Şengül (2015) categorized Google Trends (GT) data within the framework of a prediction model built using principal component analysis. The study demonstrates its superiority over other models in terms of performance by considering GT data and non-agricultural unemployment rates in its estimation model. Similarly, Bolivar et al. (2019) also tried to improve forecasting performance by using GT data on terms such as 'finding a job,' 'unemployment benefits,' and 'unemployment insurance.' The authors integrate variables such as production index, electricity consumption, capacity utilization rate, and unemployment benefit claims into their model. It is revealed that the model using GT data produces superior results compared to the models without GT data. In Şentürk (2022), where the addition of GT data to the forecasting model is advocated, the unemployment rate in Türkiye is estimated. The forecast accuracy is compared with ARIMA and ARIMAX methods.

The idea that using Google Trends (GT) data in forecasting models can significantly enhance predictive accuracy has been supported by many studies. In order to investigate whether this result is affected by the different economic structures of countries, studies conducted for many different countries, such as Fondeur and Karame (2013) in France, Adu et al. (2023) in Ghana, Askitas and Zimmermann (2009) in Germany, D'Amuri (2009) and Naccarato et al. (2018) in Italy, Anvik and Gjelstad (2010) in Norway and Mihaela (2020) in Romania are reviewed. As a result of this review, it was concluded that the use of GT data will improve the forecasting performance. Vicente et al. (2015), Gonzalez-Fernandez and Gonzalez-Velasco (2018), and Simionescu and Cifuentes-Faura (2022a, 2022b) found similar results in Spain. The studies conducted by D'Amuri and Marcucci (2010) and Ettredge et al., (2005) for the United States also support using GT data. A similar conclusion was reached in the studies conducted by Chadwick and Şengül (2015) and Şentürk (2022) for Türkiye.

The main difference in the studies examined is that the GT data are generated with keywords selected in the national language. Focusing on Spain, Gonzalez-Fernandez and Gonzalez-Velasco (2018), Mulero and Garcia-Hiernaux (2023), Simionescu and Cifuentes-Faura (2022a), Simionescu and Cifuentes-Faura (2022b), and Vicente et al. (2015) have used time series of job search queries from

Google Trends to predict unemployment. Simionescu and Cifuentes-Faura (2022a, 2022b) identified that the top keywords for job searches in Spain are “ofertas de empleo” (job vacancies), “ofertas de trabajo” (job offers), and “desempleo” (unemployment). Vicente et al. (2015) and Gonzalez-Fernandez and Gonzalez-Velasco (2018) incorporated some of these keywords to illustrate that forecasts relying on models using Google search data outperformed alternative forecasting methods for the Spanish unemployment rate. In Mulero and Garcia-Hiernaux’ (2023) paper, keywords are selected according to a different criteria procedure. They categorized the search terms into four groups: (i) queries related to leading job search platforms (such as Infojobs, Jobday, LinkedIn, etc.); (ii) searches linked to Spanish unemployment centers, whether online or physical, public or private (e.g., Employment office, SEPE, Randstad, etc.); (iii) queries related to standard job search terms (e.g., Job offers, How to Find a Job, How to Find Work, etc.); and finally, (iv) searches directly related to companies that generate the most employment in Spain (e.g., working at Inditex, Carrefour employment, Santander job opportunities).

Mihaela (2020) explained and estimated Romania's regional unemployment at the county level for 2004-2018. In addition, the Granger causality relationship between unemployment and other indicators was investigated. The findings of the study indicate that the indicators collected through Google Trends can improve the unemployment rate forecasts in Romania.

In conclusion, a detailed evaluation of the studies in the literature shows that GT data impact the success of unemployment rate forecasting models, but the forecasting accuracy depends on the keywords chosen.

3. DATA AND METHODOLOGY

This section details the data used in the unemployment estimation model and the assumptions regarding using GT data.

3.1. Data-Google Trends

In recent years, researchers have begun to use GT to measure indicators such as social behavior, attitudes, tastes, and preferences, which are difficult to measure. This preference also has some important advantages. Firstly, GT data is cost-effective and reliable. In addition, GT data eliminates the problem of data bias, as it is derived from Internet searches that the user unconsciously discloses. In addition, GT does not have the disadvantages of traditional survey measurement. These disadvantages include participant reluctance, the idea of wasting time, and laziness. The advantages of using GT data have been demonstrated in many studies (Belej, 2022; Blazquez & Domenech, 2018; Einav & Levin, 2014; Knipe et al., 2021; Rotter et al., 2021; Sherman-Morris et al., 2011; Yeh et al., 2018). When data on human and company activities are analyzed correctly, it will help reveal trends and behaviors (Blazquez & Domenech, 2018). Collecting information with GT data also has some disadvantages. While GT searches may occasionally yield unrelated results, they may also fail to capture the entirety of the relevant search (Anvik & Gjelstad, 2010). Choosing the right keywords related to the phenomenon to be researched is very important. Several studies (Askitas & Zimmermann, 2009; Simionescu & Cifuentes-Faura, 2022a) try to overcome this criticism by creating various scenarios in the research model.

GT index takes values on a scale between 0 and 100. GT is a trending index that measures how often a keyword is searched over time relative to the total search volume. The more popular the keyword is in the selected period and region, the higher it gets ranked.GT takes a zero value for queries with a low search volume and ignores repetitive searches performed by the same machine over a very short period. Searching for multiple keywords and filtering out queries with apostrophes and special characters is possible.

This study uses monthly, seasonally adjusted unemployment data (ages 15–64), the annual inflation rate, and four different GT search datasets (Jan. 2005–May 2023) to develop an unemployment forecasting model. The rationale for using inflation as an independent variable to understand unemployment is anchored on the Phillips curve. Empirical studies such as Karahan and Uslu (2018), Atgür (2020), Kırca and Canbay (2020), and Nar (2021) have shown that the Phillips Curve is valid in the Turkish context, with evidence suggesting a negative causality from inflation to unemployment. These findings highlight the relevance of inflation as a key variable in understanding and forecasting unemployment trends in Türkiye. Typically, inflation rate data becomes available more swiftly than unemployment statistics, which could make it a valuable, timely indicator for forecasting unemployment trends. Search frequency data from GT has been analyzed using four approaches (Mulero & Garcia-Hiernaux, 2023). The approaches taken for GT are given in Table 1[1].

Table 1. Definitions of Different GT Variables

Variables	Definition	Keywords
GT1	Unemployment rate + job postings	İşsizlik oranı+iş ilanları
GT2	Popular job search websites in Türkiye	İşbul net +Toptalent.co + Yenibiriş +Eleman online +Eleman net +Kariyer net
GT3	Turkish Employment Agency (state-sponsored job search platform)	İşkur+ Türkiye İş kurumu
GT4	Creating a CV + Reference letter + Interview techniques + Job application examples	CV oluşturma +Referans mektubu+Mülakat teknikleri+İş başvurusu örnekleri

[1] Although the keywords listed in Table 1 are in Turkish, they were intentionally preserved in their original form, as the study focuses on the Turkish labor market and reflects the actual search behavior of Turkish Internet users. Translating them into English would distort the nature of the data collected from Google Trends. Similar approaches have been taken in other country-specific studies (e.g., Mulero & Garcia-Hiernaux, 2023; Simionescu & Cifuentes-Faura, 2022a for Spain), where keywords were selected in the native language to ensure representativeness.

GT1 is frequently used in the literature, including the unemployment rate and the search terms for job vacancies. GT2 contains the names of the most used job posting websites in Türkiye. GT3 gives the search frequency of the government's official job search site, while GT4 includes search terms that are researched before a job search, especially for new graduates, and are more future-oriented. GT4 gives the number of searches for search terms such as creating CVs, requesting letters of reference, interview techniques, and job application examples.

Figure 1 shows the time path graphs of the unemployment and inflation rates, while Figure 2 shows the time path of Google search data for the unemployment.

Figure 1. Time Path of the Unemployment Rate and Inflation Rate

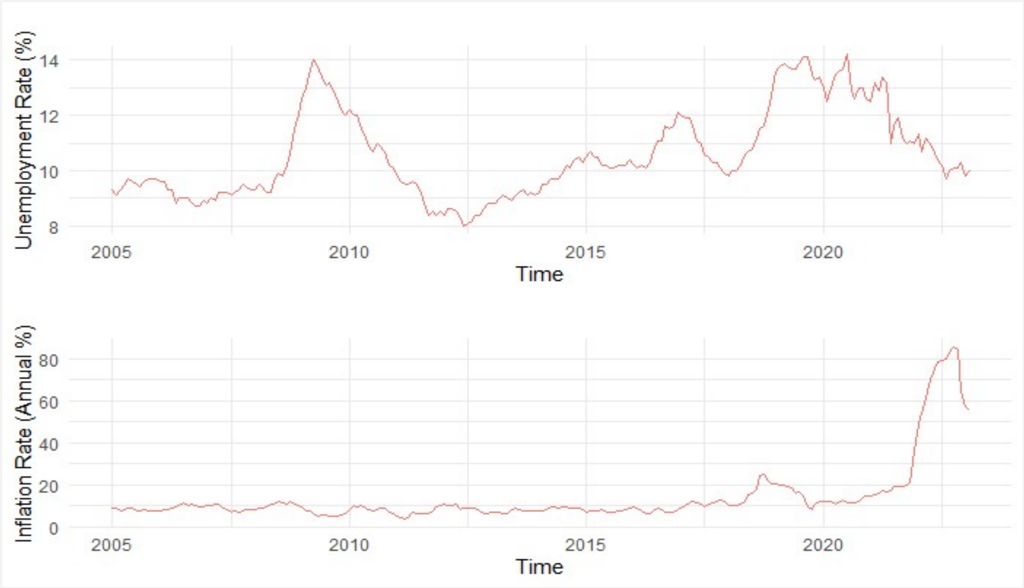
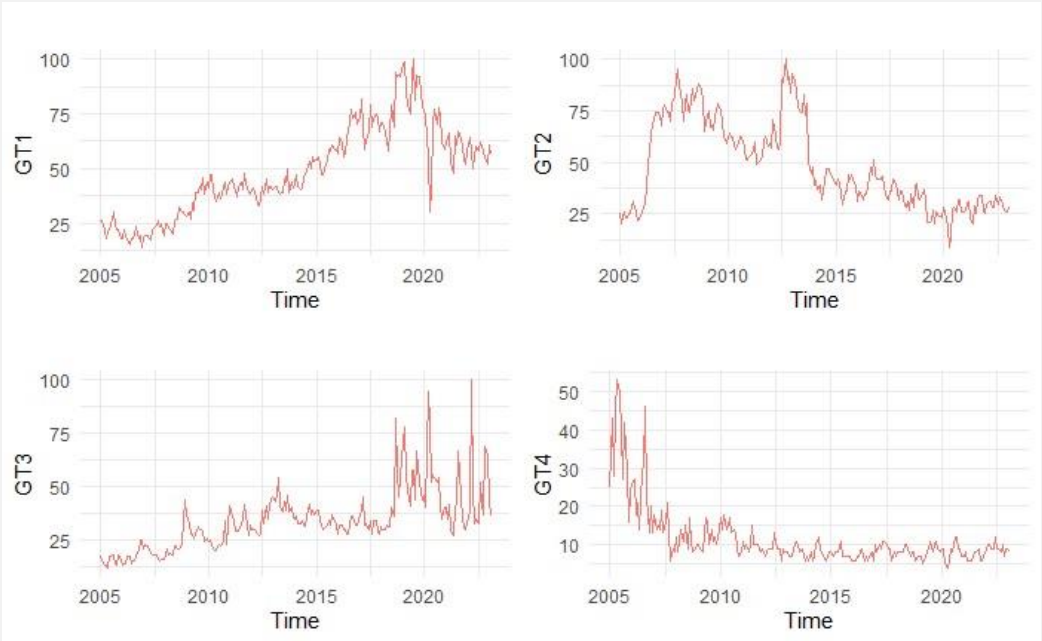


Figure 2. Time Path of Google Search Data for Unemployment



3.2. Methodology

This study considers two approaches for building a forecasting model: (1) univariate and (2) multivariate models. The univariate models used are the random walk process, ARMA, ARMAX, and SES models. In addition, VAR models are considered multivariate models.

3.2.1. Random Walk Model

Suppose that ε_t is a discrete-time purely random process with mean μ and variance σ_ε^2. A process Y_t is said to be a random walk if

$$Y_t = Y_{t-1} + \varepsilon_t$$

The random walk process is a non-stationary process. In a time series with a random walk process, future values occur randomly without depending on previous values. In such a case, predicting future values by looking at past values is impossible. However, some studies indicate that the random walk process is superior to many methods in time series forecasting (Moosa & Burns, 2014). This study considers the random walk process a benchmark model (Pesaran et al., 2009). A random walk model forecast value is defined as

$$\hat{Y}_{n+i} = Y_n \, , i = 1, 2, \dots h \, .$$

3.2.2. Simple Exponential Smoothing (SES)

Exponential smoothing (ES) is the name given to a general class of forecasting procedures that rely on simple updating equations to calculate forecasts. The most basic form introduced is simple exponential smoothing (SES), which should only be used for non-seasonal time series showing no systematic trend. Many time series that arise in practice do contain a trend or seasonal pattern, but these effects can be measured and removed to produce a stationary series for which simple ES is appropriate (Gardner Jr., 1985).

$$\hat{Y}_t = aY_t + (1-a)\hat{Y}_{t-1}$$

Here, a is the smoothing coefficient, taking values between $0 \leq a \leq 1$. Determining a is an optimization problem commonly obtained by grid search (Hyndman & Athanasopoulos, 2018).

3.2.3. Autoregressive Moving Average (ARMA)

ARMA models are among the most frequently used econometric methods for analyzing and forecasting economic time series. According to Balli and Elsamadisy (2012), the Box-Jenkins methodology is regarded as an effective prediction technique, especially for single-variable time series.

The ARMA model for a stationary time series is generally:

$$Y_t = \mu + \epsilon_t + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

ϵ_t is the white noise process, ϕ_i and θ_i are the coefficients of the autoregressive and moving average processes, respectively.

3.2.4. Autoregressive Moving Average with Explanatory Variables (ARMAX)

The ARMAX model is an extended version of the ARMA model. Unlike ARMA, exogenous (explanatory) variables are added to the model. Thanks to the added exogenous variables, the explanatory power of the model for the dependent variable increases. The mathematical representation of the ARMAX model is as follows:

$$Y_t = \mu + \epsilon_t + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \beta X_t$$

Here, X_t is the explanatory variable.

3.2.5. VAR Model

The VAR model can be considered a multivariate form of the ARMA model. Since it is a multivariate system, it is frequently used in time series analyses. In addition, unlike the ARMA model, since it is a system with more than one equation, the model's explanatory power may be higher in some cases.

Generally, VAR models is given as:

$$y_t = v + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + u_t$$

Where y_t is the variable vector, A_i's is the coefficient matrix, and u_t is the error vector.

4. EMPIRICAL RESULTS

The stationarity structure of the time series was examined prior to the empirical analysis to ensure statistically significant results from the ARMA and VAR models. Analyses with non-stationary series may be misleading due to the spurious regression problem. The Augmented Dickey-Fuller (ADF) test proposed by Dickey and Fuller (1981) is widely used to test whether the series has a unit root. However, Perron (1989) pointed out that the ADF test gives erroneous results, and the null hypothesis of unit root is not rejected in case of a break in the series. Therefore, Zivot and Andrews (1992), Perron (1997), and Vogelsang and Perron (1998) propose unit root tests that allow structural breaks to be identified endogenously from the data. The results of the unit root test are presented in Table 2.

Table 2. Unit Root Test Results

Variables	ADF Level	Break Point Unit root test (Vogelsang and Perron, 1998) Level	First Difference
ump	-2.239725 (0.193)	-2.49 (0.905	-13.45904 (< 0.01)
			Breakpoints
inf	0.115863 (0.9663)	-8.38 (< 0.01)	2020:11
gt1	-1.750679 (0.4043)	-4.22 (0.093)	2014:05
gt2	-2.114395 (0.2393)	-4.31 (0.072)	2013:07
gt3	-2.490993 (0.1191)	-7.68 (< 0.01)	2018:06
gt4	-4.221675 (0.0008)		

Note: Both trend and break features are taken as constant terms only. The break type is an additive outlier. The break selection method is the minimized Dickey-Fuller t statistic. The appropriate lag length is chosen using SIC for a maximum of 12 lags. Values in parentheses are p-values corresponding to the test statistics. Null Hypothesis: Serie has a unit root.

According to the unit root test results, the unemployment rate is the first difference stationary. Inflation, GT1, GT2, and GT3 are stationary at levels under break. The breakpoints that disrupt the stationarity structure of the series are given in Table 2. GT4 is a level stationary time series. Due to the unit root test results, the first-order difference of the unemployment series is used in the ARMA, ARMAX, and VAR models.

On the other hand, for the series that are stationary under the break, dummy variables are used for the dates of the breakpoint in the models. After these specifications, the models were estimated, and out-of-sample forecast values were obtained. For the calculation of the predictive performance of the models, the data set is divided into two parts: train and test. 75% of the dataset (163 observations) is used for model prediction, and the remaining 25% (55 observations) is used as test data for the fit of the forecast model. 2005:01 to 2018:07 is the train data set, and 2018:08 to 2023:02 is the test data set.

Numerous criteria have been put forward throughout history to assess forecast accuracy, sparking debates about their appropriate application. Root Mean Square Error (RMSE) is a standard yet scale-dependent measure, making it unsuitable for comparing models with different variables or frequencies. In this study, we used the Symmetric Mean Absolute Percentage Error (sMAPE) and Mean Absolute Scaled Error (MASE), which are better for diverse datasets. sMAPE is frequently used for its scale-independence and interpretability, while MASE mitigates potential issues from using sMAPE. The corresponding formulas are as follows:

$$\text{SMAPE} = \frac{2}{h}\sum_{t=1}^{h}\frac{2\left|Y_t - \hat{Y}_t\right|}{|Y_t| + \left|\hat{Y}_t\right|} \times 100$$

$$MASE = \frac{1}{g}\frac{\sum_{t=1}^{h}\left|Y_t - \hat{Y}_t\right|}{\frac{1}{n-m}\sum_{t=m+1}^{n}|Y_t - Y_{t-m}|}$$

Where g is the number of out-of-period predictions, $\hat{Y}_t$ is the out-of-period prediction values, m is the frequency of the data, n is the total number of observations. In the literature, many researchers have suggested using "relative" $MASE$ ($RelMASE$) and "relative" $sMAPE$ ($RelsMAPE$) values instead of using $MASE$ and $MASE$ values alone (Fildes, 1992; Ahlburg, 1992; Hyndman & Koehler, 2006). In order to calculate these criteria, it is necessary to calculate the $MAPE$ and $MAPE$ values of a simple but effective method as a performance indicator (benchmark). This study considers the Random Walk Model (RW) as a benchmark. After selecting the Random Walk Model as the performance benchmark, $RelsMAPE = sMAPE/sMAPE_{\mathrm{RW}}$ and $RelMASE = MASE/MASE_{\mathrm{RW}}$. After calculating the relative criteria, OWA criterion is calculated by taking the overall weighted average of the "relative" $MASE$ and "relative" $sMAPE$ criteria (Ağaslan & Gayaker, 2020).

Table 3. Comparison of Forecasting Performance

Models	sMAPE	MASE	RealsMAPE	RealMASE	OWA	Success Rate (%)
Random Walk	13.85677	4.58519	1	1	1	0
ARMA	16.85470	5.46424	1.21635	1.19172	1.20403	-20.40
ARMA_GT1	12.89188	4.28972	0.93037	0.93556	0.93296	6.70
ARMA_GT2	16.98718	5.50243	1.22591	1.20004	1.21298	-21.30
ARMA_GT3	16.20020	5.27857	1.16912	1.15122	1.16017	-16.02
ARMA_GT4	16.78194	5.44869	1.21110	1.18832	1.19971	-19.97
SES	13.85683	4.58521	1.00000	1.00000	1.00000	0.00
VAR	11.63288	3.89768	0.83951	0.85006	0.84478	15.52
VAR_GT1	10.89739	3.66358	0.78643	0.79900	0.79272	20.73
VAR_GT2	11.78902	3.94669	0.85078	0.86075	0.85576	14.42
VAR_GT3	10.61869	3.57825	0.76632	0.78039	0.77336	22.66
VAR_GT4	10.72192	3.61067	0.77377	0.78746	0.78062	21.94

Table 3 shows that the VAR_GT3 model exhibits the best forecasting performance. In this model, "İşkur+Turkish Employment Agency" searches are included as annual inflation and Google Trends variables. The results show that using GT data to forecast the unemployment rate improves the forecasting performance. This finding aligns with previous studies such as Chadwick and Şengül (2015) and Bolivar et al. (2019), demonstrating that integrating GT data enhances the accuracy of unemployment forecasting models. However, in univariate models, ARMA_GT1 again shows the best forecasting performance. This model is obtained using the Google Trends variable with the key term "Unemployment rate + job vacancies" in addition to the lags of the unemployment rate itself.

Similarly, Şentürk (2022) also reported that using GT variables in ARIMA and ARIMAX models significantly improves forecast accuracy for unemployment in Türkiye. Figure 3 shows out of sample forecast and actual data comparison. Figure 4 shows the out-of-sample forecast values of the VAR_GT3 model, which has the best forecasting performance.

Figure 3. Out of Sample (Ex-ante) Forecast and Actual Data

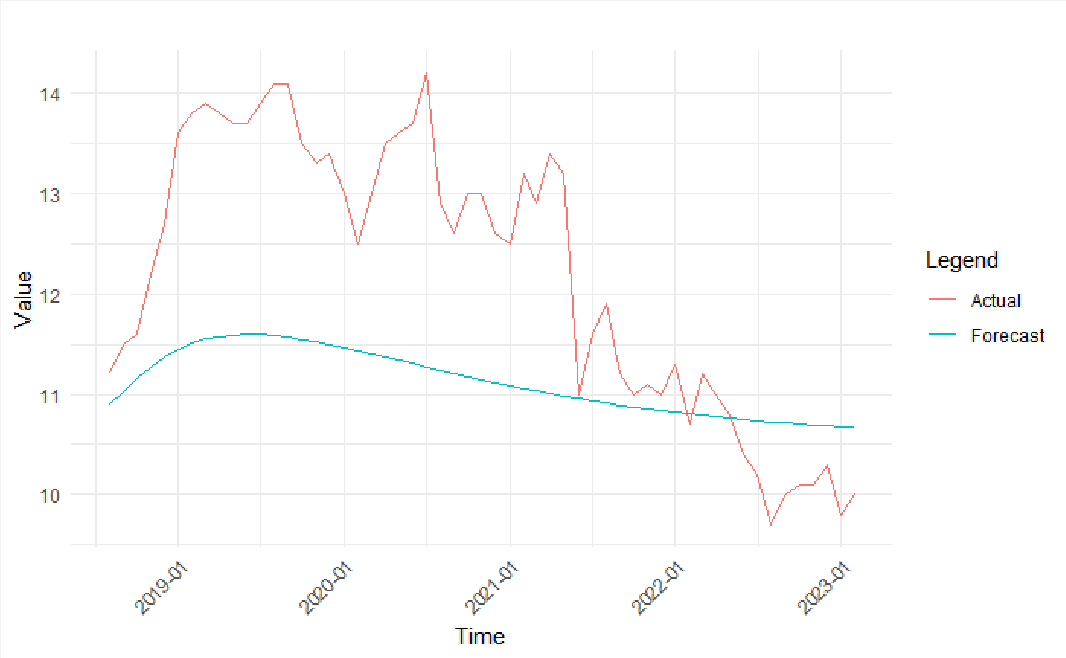
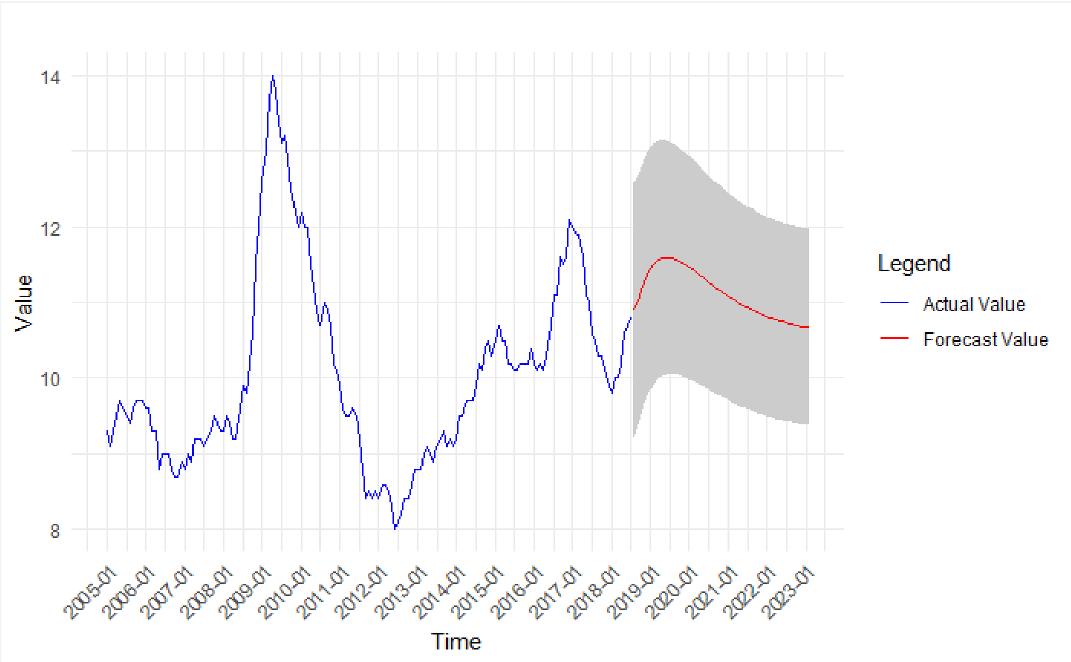


Figure 4. Actual and Out-of-Sample Forecasts with Confidence Intervals



The model results show that the models augmented with Google Trends data generally have better forecasting performance than the traditional models. The findings of this study are consistent with global literature, including Gonzalez-Fernandez and Gonzalez-Velasco (2018) for Spain and Naccarato et al. (2018) for Italy, both of which emphasize the superiority of GT-enhanced models in unemployment forecasting. In particular, the VAR_GT3 model with the annual inflation rate and search terms associated with the Turkish Employment Agency performs the best compared to the other models. This suggests that individuals' interest in state-sponsored job search platforms may be an important indicator for predicting changes in the unemployment rate. Such findings align with studies like Kırca and Canbay's (2020) and Nar (2021), highlighting the utility of inflation and unemployment causality relationships in forecasting labor market dynamics.

Moreover, the ARMA_GT1 model with Google Trends data performs best in univariate models. This model includes the key term "Unemployment rate+job openings" related to the unemployment rate. This shows that critical terms related to unemployment can be crucial in improving unemployment forecasts.

5. CONCLUSION

Economic crises, the COVID-19 pandemic, and recent waves of migration have led to a significant drop in employment, causing public concern. Türkiye faces one of Europe's highest unemployment rates, which has become a priority socioeconomic issue for the country.

This study investigates whether using Google Trends to forecast Türkiye's unemployment rate will improve the forecasting performance. For this purpose, we use seasonally adjusted unemployment rate, inflation, and four different Google Trends data between January 2005 and May 2023. The study is based on two approaches: (1) univariate models, including the random walk process, ARMA frameworks, and simple exponential smoothing techniques, and (2) multivariate models, represented by the vector autoregressive (VAR) configurations.

According to the results, the VAR_GT3 model demonstrates the highest forecasting performance. This model incorporates the annual inflation rate along with GT3, which represents Google Trends search queries related to the *Turkish Employment Agency*. The inclusion of Google Trends variables significantly enhances the accuracy of unemployment rate forecasting. In contrast, among the univariate models, ARMA_GT1 yields the best performance. This model utilizes GT1, a Google Trends variable reflecting search activity on *"unemployment rate"* and *"job postings"*, alongside the lags of the unemployment rate itself.

These findings underscore the predictive power of modern tools like Google Trends in the current digital era. Predictive models built by incorporating real-time search data into traditional models are crucial for researchers to achieve more reliable forecasting results. The superior performance of models incorporating this data type, such as VAR_GT3 and ARMA_GT1, suggests that public interest and concern - reflected in their search behavior - can predict larger economic patterns. This enhances our technical approach to forecasting and highlights the importance of considering public sentiment and behavior in economic forecasting. It is an important step in combining traditional economic metrics with data sources from the digital age.

These results hold great importance for professionals and policymakers. The impact of key terms related to unemployment, particularly on popular online platforms like Google, is essential for enhancing unemployment forecasts. As industries and policymakers try to predict job market trends, these new data sources and analysis methods can be the key to accurate and up-to-date economic forecasts.

Ethics Committee approval was not required for this study.

The authors declare that the study was conducted in accordance with research and publication ethics.

The authors confirm that no part of the study was generated, either wholly or in part, using Artificial Intelligence (AI) tools.

The authors declare that there are no financial conflicts of interest involving any institution, organization, or individual associated with this article. Additionally, there are no conflicts of interest among the authors.

The authors affirm that they contributed equally to all aspects of the research.

REFERENCES

Abraham, J., Higdon, D., Nelson, J., & Ibarra, J. (2018). Cryptocurrency price prediction using tweet volumes and sentiment analysis. *SMU Data Science Review*, *1*(3), 1.

Adu, W. K., Appiahene, P., & Afrifa, S. (2023). VAR, ARIMAX and ARIMA models for nowcasting unemployment rate in Ghana using Google Trends. *Journal of Electrical Systems and Information Technology, 10*(1), 1-16. https://doi.org/10.1186/s43067-023-00078-1

Ağaslan, E., & Gayaker, S. (2020). Türkiye ekonomisindeki emisyon hacminin düşük ve yüksek frekanslı modeller ile öngörüsü. *Bankacılık Dergisi*, *31*(114), 30-49.

Ahlburg, A. D. (1992). Predicting the job performance of managers: What do the experts know? *International Journal of Forecasting*, *7*(4), 467-472.

Anvik, C., & Gjelstad, K. (2010). " Just Google it": Forecasting Norwegian unemployment figures with web queries, *Cream*, *11*, 1–52.

Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, *55*(2), 107-120. https://doi.org/10.3790/aeq.55.2.107

Atgür, M. (2020). Inflation and unemployment relationship in Turkey: An examination on the validity of Phillips Curve (1988-2017). *International Journal of Eurasia Social Sciences*, *11*(40), 572-605. https://doi.org/10.35826/ijoess.2617

Ayyoubzadeh, S. M., Ayyoubzadeh, S. M., Zahedi, H., Ahmadi, M., & Kalhori, S. R. N. (2020). Predicting COVID-19 incidence through analysis of Google Trends data in Iran: Data mining and deep learning pilot study. *JMIR Public Health and Surveillance*, *6*(2), e18828.

Balli, F., & Elsamadisy, E. (2012). Modelling the currency in circulation for the State of Qatar. *International Journal of Islamic and Middle Eastern Finance and Management*, *5*(4), 321-339.

Belej, M. (2022). Does Google Trends show the strength of social interest as a predictor of housing price dynamics? *Sustainability, 14*(9), 5601. https://doi.org/10.3390/su14095601

Blazquez, D., & Domenech, J. (2018). Big data sources and methods for social and economic analyses. *Technological Forecasting and Social Change, 130*, 99-113. https://doi.org/10.1016/j.techfore.2017.07.027

Bolivar, F., Ortiz, A., & Rodrigo, T. (2019). Nowcasting Turkish unemployment using real-time data from Google. *BBVA Research*.

Cebrián, E., & Domenech, J. (2023). Is Google Trends a quality data source? *Applied Economics Letters, 30*(6), 811-815. https://doi.org/10.1080/13504851.2021.2023088

Chadwick, M. G., & Şengül, G. (2015). Nowcasting the unemployment rate in Turkey: Let's ask Google. *Central Bank Review*, *15*(3), 15.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record, 88*(1), 2-9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

D'Amuri, F. (2009). Predicting unemployment in short samples with Internet job search query data. *MPRA Paper 18403*, University Library of Munich, Germany.

D'Amuri, F., & Marcucci, J. (2010). 'Google it!' Forecasting the US unemployment rate with a Google job search index. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1594132

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica, 49*(4), 1057-1072. https://doi.org/10.2307/1912517

Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science, 346*(6210). https://doi.org/10.1126/science.1243089

Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM, 48*(11), 87-92. https://doi.org/10.1145/1096000.1096010

Fildes, R. (1992). Forecasting structural time series models and the Kalman filter. *International Journal of Forecasting, 8*(4), 635. https://doi.org/10.1016/0169-2070(92)90072-h

Fondeur, Y., & Karame, F. (2013). Can Google data help predict French youth unemployment? *Economic Modelling, 30*, 117-125. https://doi.org/10.1016/j.econmod.2012.07.017

Gardner Jr, E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, *4*(1), 1-28.

Gonzalez-Fernandez, M., & Gonzalez-Velasco, C. (2018). Can Google econometrics predict unemployment? Evidence from Spain. *Economics Letters*, *170*, 42-45.

Han, S. C., Chung, H., & Kang, B. H. (2012). It is time to prepare for the future: Forecasting social trends. *Communications in Computer and Information Science*, 325-331. https://doi.org/10.1007/978-3-642-35603-2_48

Hassani, H., & Silva, E. S. (2015). Forecasting with big data: A review. *Annals of Data Science, 2*(1), 5-19. https://doi.org/10.1007/s40745-015-0029-9

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*, 3rd edition, OTexts. https://otexts.com/fpp2/simple-exponential-smoothing.html

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679-688.

Jun, S. P., & Park, D. H. (2016). Consumer information search behavior and purchasing decisions: Empirical evidence from Korea. *Technological Forecasting and Social Change, 107*, 97-111. https://doi.org/10.1016/j.techfore.2016.03.021

Karahan, P., & Uslu, N. Ç. (2018). A dynamic analysis on the validity of the Phillips curve for Turkey. *Finans Politik ve Ekonomik Yorumlar*, *55*(636), 89-99.

Kırca, M., & Canbay, Ş. (2020). Kırılgan beşli ülkeler için Phillips eğrisi analizi. *İktisadi İdari ve Siyasal Araştırmalar Dergisi*, *5*(12), 130-140. https://doi.org/10.25204/iktisad.717391

Knipe, D., Gunnell, D., Evans, H., John, A., & Fancourt, D. (2021). Is Google Trends a useful tool for tracking mental and social distress during a public health emergency? A time-series analysis. *Journal of Affective Disorders, 294*, 737-744. https://doi.org/10.1016/j.jad.2021.06.086

Lee, E. (2015). Recognizing rights in real time: The role of Google in the EU right to be forgotten. *UCDL Rev.*, *49*, 1017.

McLaren, N., & Shanbhogue, R. (2011). Using Internet search data as economic indicators. *Bank of England Quarterly Bulletin,* No. 2011, Q2. http://dx.doi.org/10.2139/ssrn.1865276

Mihaela, S. (2020). Improving unemployment rate forecasts at regional level in Romania using Google Trends. *Technological Forecasting and Social Change*, *155*, 120026. https://doi.org/10.1016/j.techfore.2020.120026

Moosa, I., & Burns, K. (2014). The unbeatable random walk in exchange rate forecasting: Reality or myth? *Journal of Macroeconomics*, *40*, 69-81.

Mulero, R., & Garcia-Hiernaux, A. (2023). Forecasting unemployment with Google Trends: age, gender, and digital divide. *Empirical Economics, 65*(2), 587-605. https://doi.org/10.1007/s00181-022-02347-w

Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining official and Google Trends data to forecast the Italian youth unemployment rate. *Technological Forecasting and Social Change, 130*, 114-122. https://doi.org/10.1016/j.techfore.2017.11.022

Nar, M. (2021). Analysis of the Phillips Curve: An assessment of Turkey. *The Journal of Asian Finance. Economics, and Business*, *8*(2), 65-75.

Park, S., Lee, J., & Song, W. (2017). Short-term forecasting of Japanese tourist inflow to South Korea using Google Trends data. *Journal of Travel & Tourism Marketing*, *34*(3), 357-368.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica, 57*(6), 1361-1401. https://doi.org/10.2307/1913712

Perron, P. (1997). Further evidence on breaking trend functions in macroeconomic variables. *Journal of Econometrics*, *80*(2), 355-385. https://doi.org/10.1016/s0304-4076(97)00049-3

Pesaran, M. H., Schuermann, T., & Smith, L. V. (2009). Forecasting economic and financial variables with global VARs. *International Journal of Forecasting*, *25*(4), 642-675. https://doi.org/10.1016/j.ijforecast.2009.08.007

Rotter, D., Doebler, P., & Schmitz, F. (2021). Interests, motives, and psychological burdens in times of crisis and lockdown: Google Trends analysis to inform policymakers. *Journal of Medical Internet Research*, *23*(6), e26385. https://doi.org/10.2196/26385

Santillana, M., Nguyen, A. T., Dredze, M., Paul, M. J., Nsoesie, E. O., & Brownstein, J. S. (2015). Combining search, social media, and traditional data sources to improve influenza surveillance. *PLOS Computational Biology*, *11*(10), e1004513. https://doi.org/10.1371/journal.pcbi.1004513

Şentürk, G. (2022). Can Google search data improve the unemployment rate forecasting model? An empirical analysis for Turkey. *Journal of Economic Policy Researches*, *9*(2), 229-244.

Sherman-Morris, K., Senkbeil, J., & Carver, R. (2011). Who's Googling what? What internet searches reveal about hurricane information seeking. *Bulletin of the American Meteorological Society*, *92*(8), 975-985. https://doi.org/10.1175/2011bams3053.1

Simionescu, M. (2020). Improving unemployment rate forecasts at regional level in Romania using Google Trends. *Technological Forecasting and Social Change, 155*, 120026.

Simionescu, M., & Cifuentes-Faura, J. (2022a). Forecasting national and regional youth unemployment in Spain using Google Trends. *Social Indicators Research*, *164*(3), 1187-1216. https://doi.org/10.1007/s11205-022-02984-9

Simionescu, M., & Cifuentes-Faura, J. (2022b). Can unemployment forecasts based on Google Trends help government design better policies? An investigation based on Spain and Portugal. *Journal of Policy Modeling*, *44*(1), 1-21. https://doi.org/10.1016/j.jpolmod.2021.09.011

Simionescu, M., & Zimmermann, K. F. (2017). Big data and unemployment analysis. *GLO Discussion Paper*, No. 81. Global Labor Organization (GLO), Maastricht.

Suhoy, T. (2009). Query indices and a 2008 downturn: Israeli data, (No. 2009.06). *Bank of Israel Working Papers 2009.06*, Bank of Israel.

Vicente, M. R., López-Menéndez, A. J., & Pérez, R. (2015). Forecasting unemployment with Internet search data: Does it help to improve predictions when job destruction is skyrocketing? *Technological Forecasting and Social Change*, *92*, 132-139. https://doi.org/10.1016/j.techfore.2014.12.005

Vogelsang, T. J., & Perron, P. (1998). Additional tests for a unit root allowing for a break in the trend function at an unknown time. *International Economic Review*, *39*(4), 1073-1100. https://doi.org/10.2307/2527353

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

Wilcoxson, J., Follett, L., & Severe, S. (2020). Forecasting foreign exchange markets using Google Trends: Prediction performance of competing models. *Journal of Behavioral Finance*, *21*(4), 412-422.

Wu, L., & Brynjolfsson, E. (2015). The future of prediction: How Google searches foreshadow housing prices and sales. In A. Goldfarb, S. M. Greenstein & C.E. Tucker (Eds.), *Economic Analysis of the Digital Economy* (pp. 89-118). University of Chicago Press.

Yeh, C. H., Wang, Y. S., Lin, S. J., Tseng, T. H., Lin, H. H., Shih, Y. W., & Lai, Y. H. (2018). What drives internet users' willingness to provide personal information? *Online Information Review*, 42(6), 923–939. https://doi.org/10.1108/OIR-09-2016-0264

Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics, 10*(3), 251-270. https://doi.org/10.1080/07350015.1992.10509904