Is Google Trends Useful in Nowcasting Unemployment Rate During the Pandemic at Regional and National Level in Romania?

Mihaela Simionescu, PhD Professor (Full)
(mihaela.simionescu@ipe.ro, mihaela_mb1@yahoo.com)
Institute for Economic Forecasting, Romanian Academy

ABSTRACT

Given the role of Covid-19 pandemic in accelerating the digital transformation in Europe, the main aim of this paper is to explain the unemployment rate in Romania based on Internet searches for jobs during the epidemic at national and county level. Google Trends indexes for certain keywords related to jobs in Romanian language (”locuri de muncă” and ”joburi”) and the most famous websites with job announcements (”eJobs” and ”Hipo”) are considered. At national level, the unemployment rate in the period February 2020- December 2022 in Romania is explained using an autoregressive distributed lag model (ARDL) based on Google Trends index for ”locuri de munca”, while for youth unemployment ”joburi” and ”eJobs” are relevant. At county level, a spatial error model based on searches for ” locuri de munca” performs better than OLS regression. The results support the recommendations to improve governmental making-decision process.

Key-words: *unemployment; Google Trends; OLS; spatial model; ARDL model*

1. INTRODUCTION

The Covid-19 pandemic has accelerated digital transformation at global level, but it has also enhanced the concerns related to health and unemployment. Mangono et al. (2021) concluded that Google searches on unemployment were among the most popular searches during the pandemic in the US together with searches related to health issues. The same conclusion is suggested by Sotis (2021) for all US countries, for the District of Columbia and three US states (Nevada, Mississippi, and Utah). Using a Lotka–Volterra model, the author has proven that during the pandemic the relationship between Google searches on unemployment, news and symptoms has become strong.

Google Trends accessed through https://trends.google.com/trends/ provides information related to requests made to the Google search engine in all the countries. This tool is based on a random sample representative of all queries which are handled by Google daily (Caperna et al., 2020). Normalization of searches is made based on location of the request and time. Search results are normalized to the time and location of a query. Each data point for a specific time and location is divided by the sum of all searches to get the relative popularity that is scaled on a range of 0 to 100 considering one query's proportion to total searches on total queries. The index is named Google Trends Index and is denoted by GTI. Two main advantages are explained by Caperna et al. (2022): provision of data for indicators that are released with low frequency from official sources and the low sensitiveness to the small sample bias.

Even if it is a powerful tool, Google Trends presents more limitations. It depends on the Internet penetration, social and economic status of the individuals, and category of age. Young people are more likely to use Google to search for jobs. On the other hand, the old people are likely to be retired and are not interested in using Internet to find out job opportunities. People with high degree of poverty and/or illiterate that are looking for a job have less chances to use the benefits of technological progress to integrate on labour market.

The research question is related to the utility of Google in explaining the unemployment in real time in Romania during the pandemic when digital transformation has accelerated. The hypothesis that will be checked is that Internet searches for jobs explained unemployment rate in pandemic at national and regional level. Considering this hypothesis, the main objective of this paper is the evaluation of the capacity of Google searches for jobs to explain and predict unemployment rate. Starting from this general objective, two specific objectives are analyzed based on a regional approach: the hypothesis is checked at national level and at county level. Therefore, the time series approach at national level (OLS regressions and ARDL model) is combined to spatial approach at county level (OLS regressions and spatial autoregressive model). The results show common conclusions as well as differences: Google searches for the same key-words ("locuri de munca" and "joburi") explain unemployment at both levels and other key-words are relevant only at national level ("eJobs" and "Hipo").

These results are a step forward in nowcasting unemployment rate in Romania, a country where official statistics are still released late. The paper continues with a short presentation of previous achievements in different countries on the unemployment- Google searches nexus. The next sections present methodology, results with discussion and conclusions.

2. LITERATURE REVIEW

There are two major international contexts that have intensified the research around the utility of Google Trends in nowcasting and forecasting unemployment rate. In the first stage, the use of Google Trends is related to the necessity to anticipate the evolution on labour market in the context of global economic crisis. Therefore, econometric models were built mostly on time series to explain the evolution of unemployment in real time. Besides the econometric models on time series, some studies checked for Granger causality between unemployment and Google searches for jobs (Askitas, Zimmermann 2009a, Su 2014) and the causality was validated in one way or another. The Granger causality test on stationary time series makes also the subject of our paper. Two seminal papers of Askitas and Zimmermann (2009 a,b) established a significant relationship between unemployment and Google searches for jobs in Germany. The key-words used for searches refer to unemployment state (*unemployment rate*), job opportunities (*job search, short-term work most popular search engines in Germany*) and various institutions that manage the unemployed issues (*labour office, unemployment office or agency, Personnel Consultant*). In our paper, we employed as key-words those expressions referring to job opportunities (locuri de munca, joburi- words used for jobs in Romanian language) and the most important websites used for job searches (eJobs and Hipo). A previous study for Romania of Simionescu (2020) used two of our key-words (locuri de munca and joburi) and another similar word (angajari). However, the paper of Simionescu (2020) did not take into account the websites from Romania where job offers are posted. Two arguments are provided for these types of key-words in our paper. First, job opportunities are more appropriate than unemployment, since someone who is no unemployed might search for unemployment to understand better the concept. Second, websites used for job searches are more relevant than labour offices during the pandemic when people try to limit the physical contact as much as possible. Moreover, the efforts for digital transformation during the epidemic made online search a more simple and affordable way to search for jobs.

In general, the digital transformation has been made fast in developed countries even before pandemic. Therefore, in pre-pandemic period most of the studies discussing the role of Google Trends in anticipating the unemployment were made for developed countries like Norway, UK, Italy, France, Spain, Canada, US, Israel (Anvik, Gjelstad 2010, McLaren, Shanbhogue 2011, D'Amuri 2009, Naccarato et al. 2018; Fondeur, Karamé 2013, Vicente et al. 2015, Dilmaghani 2019, Choi, Varian 2012, Suhoy 2009). However, Google Trends was not a successful tool in improving the unemployment rate forecasts in all countries. For example, Barreira et al. (2013) indicated that only the predictions for France, Italy, and Portugal improved by including Google Trends indexes, while in Spain this tool was not efficient.

Only few studies addressed this topic in developing countries like Brazil, Turkey, Ukraine, V4 countries (Hungary, Poland, Czech Republic, Slovakia) and Romania (Lasso, Snijders 2016, Chadwick, Sengül 2015, Pavlicek, Kristoufek 2015, Oleksandr 2010, Simionescu 2020), because of low Internet penetration rate. Only the predictions made for Hungary and Czechia on the horizon January 2004 - December 2013 improved by taking into account Google queries for jobs, while in Poland and Slovakia the correlation between unemployment and Google Trends indexes was not significant (Pavlicek, Kristoufek 2015). For Romania, in pre-pandemic period, Simionescu (2020) showed that quarterly unemployment forecasts at county level based on dynamic panel data model and Google queries related to jobs are better than predictions based on models that do not include Internet data.

In the second stage, many papers analyzing the connection between unemployment and Google Trends have appeared in the context of Covid-19 pandemic and our paper belongs to this strand. Therefore, a deeper presentation is made to the research direction related to Google Trends and unemployment during the Covid-19 epidemic. Doerr and Gambacorta (2020) showed that the US regions that were more affected by Covid-19 pandemic reported more Google searches related to unemployment and pandemic.

Most of the papers implement a time series approach or panel data models and analyzed one specific country or a sample of countries. Few recent studies analyzed only one country using time series models. For example, for India the authors Fajar et al. (2020) predicted unemployment rate in Indonesia in the first months of pandemic (March-June 2020) using an ARIMAX model and "phk" (work termination) as key-word. Yurevich and Akhmadeev (2021) predicted unemployment rate in Russia during the pandemic using autoregressive models and a hybrid model that include Google Trends indexes related to job search. The NEET unemployment in Italy was nowcasted and forecasted for few years (2019-2021) by Fenga and Son-Turan (2020) using a feed-forward artificial neural network that includes Google Trend index related to job searches. Yi et al. (2021) proposed a semiparametric method called Penalized Regression with Inferred Seasonality Module based on Google Trends data to predict unemployment during the pandemic in the US.

A SARIMA model (Seasonal Autoregressive Integrated Moving Average) was used by Roopnarine and Spencer (2021) to forecast unemployment in Trinidad and Tobago.

There are other studies made for a sample of countries. A complex procedure is described by Caperna et al. (2022) for the EU-27 countries. The authors retrieved more than 400 queries connected to unemployment in each national language. The selection of the best queries is based on machine learning techniques that allow the combination of the queries and the creation of specific indicators related to these searches. Borup and Schütte (2022) showed that a panel data approach based on Google Trends information outperformed a time series approach to predict unemployment in US countries. Moreover, Brave et al. (2020) showed a positive connection between searches on unemployment and the rate of unemployment insurance during the pandemic in the US, which is explained by the variation across time for metro areas and less by variation in space. There are significant differences in this correlation during the last global economic crisis and during epidemic because of the federal Pandemic Unemployment Assistance (PUA) program.

The cross-sectional data are rarely used in papers related to Google Trends and pandemic. For example, Larson and Sinclair (2022) showed that cross-sectional data based on Google searches for jobs in the US states include relevant information that could improve forecasts of unemployment claims in spring 2020, but on longer periods a time series approach based on autoregressive models is better.

In this paper, the approach based on time series at national level has been combined with that based on cross-sectional data at county level to explain the monthly and the annual unemployment, respectively during the pandemic. The aim is not to make predictions, but to understand better if Google searches for specific key-words explained the unemployment rate during the pandemic. In this context, there are high chances to nowcast with enough accuracy the evolution of the unemployment rate in a dynamic framework based on digital transformation.

3. METHODOLOGY AND DATA

The methodology corresponds to the two specific objectives of the paper that responds to the research question at two levels: national level and county level. The analysis from a national perspective is based on time series and employs an ARDL model and OLS regressions. The research at county level is based on OLS regressions and spatial autoregressive models developed on cross-sectional data for 2020 and 2021, respectively, years corresponding to Covid-19 pandemic.

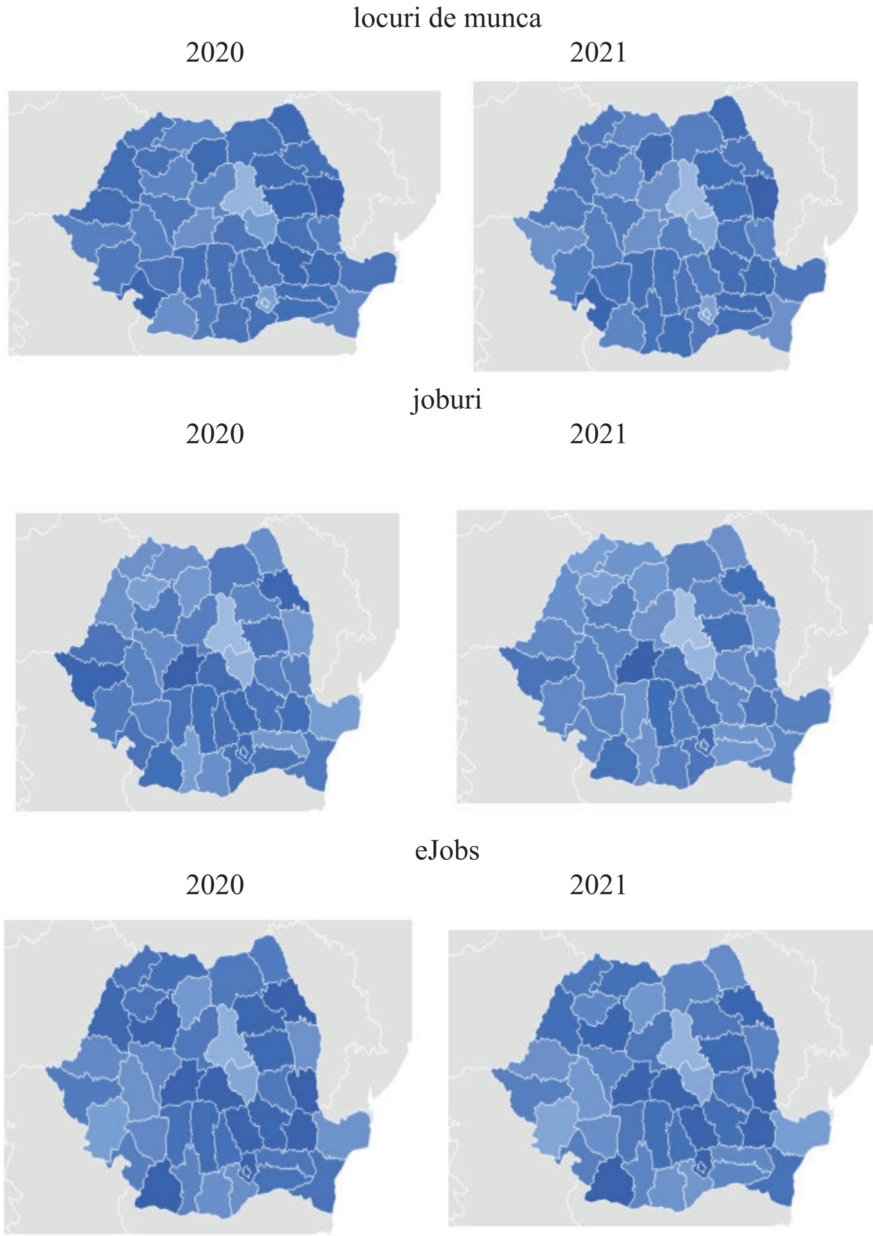
At national level, the data frequency is monthly starting with the first month of Covid-19 pandemic (February 2020) until December 2022. The unemployment rate (15-74 years) provided by Tempo online (official database of the National Institute of Statistics in Romania) and youth unemployment rate (15-24 years) from Eurostat are used as dependent variables in the models. The explanatory variables are represented by inflation rate based on index of consumer prices (monthly average value) denoted by inflation, real average wage (the nominal value is adjusted for inflation) denoted by wage, and Google Trend indexes for key-words related to job searching: ”locuri de munca” and ”joburi” that mean jobs and ”eJobs” and ”Hipo” that are two websites with job offers. The matrix of correlation for explanatory variables indicates no strong coefficients of correlation (values under 0.35).

At county level, the analysis is conducted for two different years covering the pandemic: 2020 and 2021. The annual data series for unemployment rate used as dependent variable is taken from Tempo online. The explanatory variables with annual data series correspond to more indicators with no significant correlations between them in each year: number of emigrants in each county (permanent and temporary migrants) provided by Tempo online (underestimated values, but more plausible compared to pre-pandemic values), real annual average wage denoted by wage, and monthly Google Trend indexes for key-words related to job searching: ”locuri de munca”, ”joburi”, ”eJobs” and ”Hipo”.

There are significant differences between the distributions of the counties according to searches for various key-words. According to Figure 1, the highest searches for ”locuri de munca” in 2020 and 2021 were made in Vaslui county (maximum value in 2020 and 2021) located in the Eastern of Romania and Mehedinti county in the South-West of the state. Sibiu (center of the country) was the county with maximum searches after ”joburi” in 2020 and 2021. The most of the searches after ”eJobs” were made in Iasi in 2020 (Eastern region) and in Dolj (South-Oltenia region) in 2021. In many counties the queries for ”Hipo” were not significant.

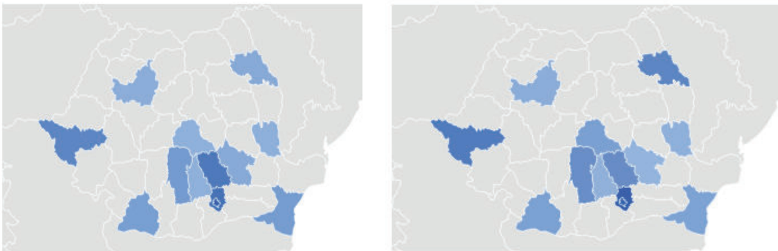
Google Trends indexes associated to searches for various key-words

Figure 1



Hipo

2020 2021



Source: Google Trends

The time series approach starts from a basic model that is extended by adding control variables:

$$unemployment_t = \alpha_{01} + \alpha_{11} \cdot GTI_t + \varepsilon_{1t}$$

$$unemployment_t = \alpha_{02} + \alpha_{12} \cdot GTI_t + \alpha_{22} \cdot wage_t + \alpha_{32} \cdot inflation_t + \varepsilon_{2t}$$

t- index for month

unemployment- dependent variable represented by total unemployment rate or youth unemployment rate

GTI- Google Trends index for a specific key-word ”locuri de munca”/ ”joburi”/”eJobs”/”Hipo”.

wage- monthly average real wage

inflation- inflation rate based on index of consumer prices

$\alpha_{01}, \alpha_{11}, \alpha_{02}, \alpha_{12}, \alpha_{22}, \alpha_{32}$- parameters to be estimated using the data series and ordinary least squares as estimation method

$\varepsilon_{1t}, \varepsilon_{2t}$- error terms

An autoregressive distributed lag model (ARDL) is also built. This type of model includes among regressors the dependent variable with a certain lag and lagged explanatory variables. Prior to estimations, a unit rot test should be applied to check for the order of cointegration. The results in the next section will indicate that GTI associated to all key-words are stationary in level, while the rest of the data series are integrated of order one (unemployment rate, youth unemployment rate, inflation rate, and wage) at 1% significance level. Given the fact that some data series are I(0) and others are I(1), an ARDL model could be constructed.

The ARDL models present more advantages: the description of short-run and long-run relationship when the time series are integrated of different orders, but without an order higher than 1 and superior estimations for small samples. Multicollinearity and serial correlation of errors might be potential issues when OLS is used. If Y is the endogenous variable and X_1, X_2,..., X_k represent the k explanatory variables, the general form of the ARDL (p, q_1, q_2,..., q_k) model is:

$$Y_t = \alpha_0 + \alpha_1 t + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{j=1}^{k} \sum_{l_j=0}^{q_j} \gamma_{j,l_j} X_{j,t-l_j} + \varepsilon_t \quad (1)$$

t- trend

α_0- intercept

α_1- parameter associated to trend

β_i- parameters associated to lagged values of Y

γ_{j,l_j}- parameters associated to lagged values of the k explanatory variables for j=1, 2,..., k

ε_t- innovations

If L is the lag operator, the polynomials in L are denoted by $\beta(L)$ and $\gamma_j(L)$:

$$\beta(L) = 1 - \sum_{i=1}^{p} \beta_i L^i$$

$$\gamma_j(L) = 1 - \sum_{l_j=1}^{q_j} \gamma_{j,l_j} L^{l_j}$$

Considering the polynomials in L defined above the equation (1) becomes:

$$\beta(L) Y_t = \alpha_0 + \alpha_1 t + \sum_{j=1}^{k} \gamma_j(L) X_{j,t} + \varepsilon_t \quad (2)$$

The analysis at county level is based on OLS regressions and spatial regressions to account for any spatial autocorrelation. The basic models and the extended one are represented below:

$$unemployment_i = \beta_{01} + \beta_{11} \cdot GTI_i + u_{1i}$$

$$unemployment_i = \beta_{02} + \beta_{12} \cdot GTI_i + \beta_{22} \cdot wage_i + \beta_{32} \cdot emigrants_i + u_{2i}$$

i-index for county

$\beta_{01}, \beta_{11}, \beta_{02}, \beta_{12}, \beta_{22}, \beta_{32}$- parameters to be estimated using the data series and OLS

u_{1i}, u_{2i} - error terms

Spatial autocorrelation should be taken into account since neighbouring counties might have similar values for unemployment. Moran's I statistic is used to check for spatial autocorrelation. If spatial dependence is accepted, then a spatial model could be constructed. One type of model is the spatial autoregressive model (SAR) with a spatial lag associated to dependent variable:

$$unemployment_t = \rho W unemployment_i + \gamma_{11} \cdot GTI_t + \gamma_{21} \cdot wage_t + \gamma_{31} \cdot emigrants_i + v_{1i}$$

The spatial error model (SEM) is another option that considers spatial dependence in the error:

$$v_{1i} = \lambda W v_{1i} + w_i$$

$$unemployment_t = \gamma_{02} + \gamma_{12} \cdot GTI_t + \gamma_{22} \cdot wage_t + \gamma_{32} \cdot emigrants_i + \lambda W v_{1i} + w_i$$

Lagrange multiplier test for spatial error and lag is used to select the best model. If the null hypothesis (spatial randomness) is not rejected, the OLS is better than the spatial model.

4. RESULTS AND DISCUSSION

At national level, a time series approach is employed based on ARDL model and OLS regression for the period February 2020-December 2022. All the data series for the variables mentioned in the previous section are seasonally adjusted using Tramo/Seats method. The Augmented Dickey-Fuller (ADF) is applied to check for unit root in the seasonally adjusted time series. The ADF test to check for unit root was considered following the descending strategy of Dickey and Pantula. In all the cases, the most suitable model for ADF equation included no trend and no intercept. The results in Table 1 indicated that unemployment rate, youth unemployment rate, inflation rate and wage are integrated of order 1, while the data series corresponding to different Google queries are stationary in level at 1% significance level.

The results of ADF test

Table 1

Variable	Type of data series	ADF stat.	p-value
unemployment rate	Data series in the second difference	10.015	<0.01
	Data series in the first difference	-5.190	<0.01
	Data series in level	0.266	0.757
youth unemployment rate	Data series in the second difference	-9.646	<0.01
	Data series in the first difference	-5.654	<0.01
	Data series in level	0.397	0.792
inflation rate	Data series in the second difference	-11.092	<0.01
	Data series in the first difference	-7.570	<0.01
	Data series in level	0.231	0.747
wage	Data series in the second difference	-9.777	<0.01
	Data series in the first difference	-5.320	<0.01
	Data series in level	3.342	0.999
locuri de munca	Data series in the second difference	-5.782	<0.01
	Data series in the first difference	-7.572	<0.01
	Data series in level	-6.608	<0.01
Joburi	Data series in the second difference	-5.399	<0.01
	Data series in the first difference	-5.109	<0.01
	Data series in level	-5.620	<0.01
eJobs	Data series in the second difference	-5.436	<0.01
	Data series in the first difference	-5.151	<0.01
	Data series in level	-5.655	<0.01
Hipo	Data series in the second difference	-6.215	<0.01
	Data series in the first difference	-7.033	<0.01
	Data series in level	-6.442	<0.01

Source: own calculations in EViews.

Granger causality could be checked only on stationary data series. According to Table 2, the variation in unemployment rate is cause for all the key-words excepting ”locuri de munca”, at different significance levels (1% for ”joburi”, 5% for ”eJobs” and 10% for ”Hipo”), but the reciprocal causality is not checked. On the other hand, the variation in youth unemployment rate is Granger cause only for GTI related to ”joburi” and ”eJobs” and none of the Google Trends indexes are causes for variation in youth unemployment.

The results of Granger causality test

Table 2

Cause	Effect	Stat.	p-value
Δunemployment	locuri de munca	1.790	0.102
Δunemployment	Joburi	5.496	0.0097
Δunemployment	eJobs	4.043	0.0287
Δunemployment	Hipo	2.681	0.0860
locuri de munca	Δunemployment	2.503	0.1000
Joburi	Δunemployment	1.229	0.3077
eJobs	Δunemployment	1.034	0.3686
Hipo	Δunemployment	0.955	0.3967
Δ youth unemployment	locuri de munca	0.014	0.9857
Δ youth unemployment	Joburi	5.802	0.0078
Δ youth unemployment	eJobs	4.125	0.0225
Δ youth unemployment	Hipo	0.187	0.8298
locuri de munca	Δ youth unemployment	0.109	0.8966
Joburi	Δ youth unemployment	0.178	0.8375
eJobs	Δ youth unemployment	0.868	0.4305
Hipo	Δ youth unemployment	1.706	0.1998

*Source: own calculations in EViews. Note: * for p-value<0.1, ** for p-value<0.05, *** for p-value<0.01*

Since unemployment series is integrated of order one (I(1)) and is not cause for ”locuri de munca”, inflation and wage, an ARDL model might be considered in this case. An ARDL(1,2,2,1) model was built to explain registered unemployment rate and only searches in the previous one and two months had impact on unemployment. The diagnostic tests suggest that the errors are independent, homoskedatic and normally distributed, while the model is correctly specified.

The Table 3 indicates that only after two months of searches for jobs, the unemployment at national level begins to reduce. Wage in the previous two months had a negative effect on unemployment rate which supports the idea that the increase in wage in the long run motivates unemployed to get hire faster. Inflation has a small and negative impact on registered unemployment rate. The rest of the key-words were not relevant in explaining unemployment at national level in the period February 2020-December 2022.

ARDL(1,2,2,1) model to explain registered unemployment rate in the actual month in Romania based on Google searches for ”locuri de munca” (jobs in Romanian language)

Table 3

Variable	Coefficient	p-value
unemployment(t-1)	0.836***	0.0000
locuri de munca (t)	-0.004	0.7547
locuri de munca (t-1)	0.05***	0.0062
locuri de munca (t-2)	-0.029**	0.0480
wage(t)	0.0001	0.3197
wage(t-1)	0.0003	0.1537
wage(t-2)	-0.0004**	0.0279
inflation(t)	-0.041*	0.0943
inflation(t-1)	-0.046*	0.0676
constant	9.037***	0.0063
Diagnostic tests		
R-square	0.8836	-
DW test (stat.)	1.931	-
Breusch-Godfrey Serial Correlation LM Test (autocorrelation of order 2) (stat. & p-value)	0.190	0.909
Breusch-Pagan-Godfrey test (stat. & p-value)	9.398	0.4013
Jarque-Bera test (stat. & p-value)	0.388	0.823
Ramsey RESET Test (stat. & p-value)	0.870	0.393

*Source: own calculations in EViews. Note: * for p-value<0.1, ** for p-value<0.05, *** for p-value<0.01*

In the case of variation in monthly youth unemployment rate, only the GTI for ”joburi” and ”eJobs” had a positive and significant impact on the dependent variables (see Table 4). ”Joburi” is the English takeover of the expression ”locuri de munca” that is more popular among young people. Moreover, ”eJobs” is the largest online recruitment platform in Romania, with an average of two million daily users. Variation in inflation had no effect on changes in youth unemployment rate, while the increase of wage from one month to another reduced the unemployment, but the influence is very small. This supports the hypothesis that young people are more motivated to get a job due to salary (Buheji, 2019). Actually, the salary is a target for young people because it offers to young people a financial independence from their parents (Bea and Yi, 2019).

OLS regressions to explain the variation in monthly youth unemployment rate in Romania using Internet searches for jobs

Table 4

Variable	Coefficient (p-value in brackets)			
Δwage(t)	-0.002*** (0.0006)	-0.002*** (0.0006)	-0.002*** (0.0005)	-0.0025*** (0.0003)
Δinflation(t)	0.097 (0.677)	0.124 (0.594)	0.132 (0.570)	0.117 (0.611)
locuri de munca (t)	0.0067 (0.413)	-	-	-
joburi (t)	-	0.013** (0.076)	-	-
eJobs (t)	-	-	0.018** (0.073)	-
Hipo(t)	-	-	-	0.0031 (0.625)
constant	2.361 (0.9162)	-0.873 (0.969)	-1.440 (0.948)	-0.608 (0.978)
Diagnostic tests				
R-square	0.426	0.423	0.425	0.434
DW test (stat.)	1.894	1.891	1.896	1.902
Breusch-Godfrey Serial Correlation LM Test (autocorrelation of order 2) (stat. & p-value)	0.224 (0.899)	0.187 (0.903)	0.229 (0.900)	0.195 (0.901)
Breusch-Pagan-Godfrey test (stat. & p-value)	5.275 (0.152)	3.194 (0.389)	3.679 (0.298)	1.846 (0.605)
Jarque-Bera test (stat. & p-value)	1.201 (0.548)	3.472 (0.180)	2.829 (0.242)	3.461 (0.177)
Ramsey RESET Test (stat. & p-value)	1.339 (0.191)	1.398 (0.172)	1.947 (0.173)	0.152 (0.699)

*Source: own calculations in MATLAB. Note: * for p-value<0.1, ** for p-value<0.05, *** for p-value<0.01*

The approach based on cross-sectional data for 2020 and 2021 (years of pandemic) is based on OLS regressions and spatial models. There is a negative, but very small influence of wage and number of emigrants on unemployment in 2020 and 2021 as Table 5 indicates. Google searches for ”locuri de munca” has a positive, but very low impact on unemployment rate. On the other hand, OLS suggested a positive effect of searches for ”joburi” on unemployment. The SAR and SEM models based on GTI for ”joburi” indicated that the coefficient for this variable is not statistically significant. More searches for ”locuri de munca” indicate higher unemployment rate at county level.

Moran’s I value is 2.933 with p-value less than 0.01, which indicates spatial correlation for unemployment rate. Consequently, the estimations for the entire country could not properly explain the unemployment in any county. The existence of spatial dependency implies the utilisation of spatial models (SAR or SEM) as Anselin (2005) indicated. The parameters of these models were estimated using maximum likelihood method. The results of likelihood ratio test suggest that OLS model is better than spatial lag model (p-value higher than 0.05). On the other hand, the parameters for spatial lag variables are not significant at 5% level. Having the spatial lag model rejected, the conclusion is that unemployment in a county was not correlated with the unemployment from neighbouring counties.

According to likelihood ratio test, the spatial error model performs better than OLS model. Moreover, the parameter of lambda is statistically significant at 5% level. Therefore, we can conclude that the spatial error model explains better the unemployment rate in Romania at county level in 2020 and 2021. Therefore, the spatial dependence in unemployment is present among Romanian counties, but there are other variables that are not included in this model that explain the correlation between neighbouring counties.

OLS regressions and spatial models to explain monthly unemployment rate in Romania using Internet searches for jobs during the Covid-19 pandemic (in 2020 and 2021)

Table 5

Variable	Coefficient (p-value in brackets)								
	2020	2020	2020	2020	2020	2021	2021	2021	2021
W unemplyoment (spatial lag)	-	-	-	0.225 (0.112)	-	-	0.187 (0.128)	-	-
λ	-	-	-	-	0.432** (0.016)	-	-	-	0.371** (0.022)
wage(t)		-0.003*** (0.0055)	-	-0.002*** (0.005)	-0.001*** (0.006)	-0.001***(0.007)	-	-	-0.001*** (0.007)
emigrants(t)	-	0.006** (0.012)	-	0.004** (0.022)	0.0043 (0.021)	0.003** (0.024)	-	-	0.0033** (0.024)
locuri de munca (t)	0.038** (0.014)	0.056*** (0.0004)	-	0.044*** (0.0009)	0.045*** (0.0009)	0.054*** (0.0009)	0.055*** (0.0009)	-	0.061*** (0.0006)
joburi (t)	-	-	-0.029** (0.036)	-	-	-	-	-0.030* (0.052)	-
constant	1.329 (0.212)	-1.932 (0.1479)	5.752*** (0.0001)	4.765*** (0.001)	2.233** (0.011)	0.199 (0.843)	0.225 (0.769)	5.369*** (0.000)	2.445** (0.023)
Diagnostic tests									
R-square	0.439	0.576	0.487	0.606	0.678	0.475	0.511	0.433	0.536
DW test (stat.)	2.053	2.315	1.957	-	-	1.776	-	1.899	-
Breusch-Godfrey Serial Correlation LM Test (autocorrelation of order 2) (stat. & p-value)	0.719 (0.697)	1.115 (0.5725)	0.623 (0.424)	-	-	0.603 (0.739)	-	2.191 (0.334)	-
Breusch-Pagan test (stat. & p-value)	4.327 (0.114)	7.963 (0.537)	2.580 (0.275)	3.199 (0.196)	4.022 (0.181)	3.341 (0.188)	3.556 (0.136)	1.080 (0.582)	3.2 (0.228)
Jarque-Bera test (stat. & p-value)	3.759 (0.152)	3.118 (0.210)	1.372 (0.503)	-	-	4.585 (0.1001)	-	4.570 (0.101)	-
Ramsey RESET Test (stat. & p-value)	0.655 (0.485)	0.584 (0.562)	3.632 (0.162)	-	-	3.332 (0.202)	-	0.652 (0.424)	-

Likelihood Ratio test (stat. & p-value)	-	-	-	2.553 (0.117)	4.022 (0.024)	-	2.422 (0.123)	-	3.96 (0.03)
Log likelihood	-355.27	-356.67	-352.49	-340.18	-338.41	-390.48	-370.36	-379.34	-369.46

*Source: own calculations in MATLAB. Note: * for p-value<0.1, ** for p-value<0.05, *** for p-value<0.01*

These findings are subject to discussion. At county level, higher average wage motivates the unemployed to search more for a job, while the emigrants enhance the tensions on labour market even if from theoretical point of view emigration should make available more jobs for the unemployed that remain in the county. However, the influence is very small. The results are in line with Škuflić, and Vučković (2018) who show that emigration increased unemployment in nine EU Member States, including Romania, in the period 2004-2015. These findings might be explained by structural issues of the labour market determined by high emigration, including a significant supply and demand mismatch on the labour market. Remittances from relatives working in a foreign country might encourage unemployed to remain in this state. The positive correlation between emigration and unemployment is explained by endogenous growth theories that consider low substitutability and human capital externalities for well skilled and less skilled workers. On the other hand, the skilled emigrants can play the role the citizens that support the better quality of institutions.

5. CONCLUSIONS

Google Trends represents a powerful tool in a world where digital transformation is a priority for the EU countries. The Covid-19 pandemic has enhanced the digital transformation, but also increase the concerns related to unemployment. In this context, this paper proposes a deep analysis of the capacity of Internet searches for jobs to nowcast unemployment in Romania, a country that has made progress in digitization, but still encounters difficulties in releasing official statistics in time. This research came up with a national and also with a regional approach to make a comparative analysis. At national level, only the keyword ”locuri de munca” explained the registered unemployment rate since the beginning of the pandemic until the end of 2022 showing that it is required more time to search for jobs until unemployment decreases. On contrary, the youth unemployment rate is explained by searches after ”joburi” and ”eJobs” platform. These findings can improve decision-making process at governmental level and can ensure the best policies to reduce unemployment and create job opportunities.

At county level, ”locuri de munca” also explains the unemployment rate in 2020 and 2021 according to a spatial error model, but the spatial dependence is attributed to other factors that are not included in the model.

Besides the value of these results for Romanian labour market, this study is subject to more limitations. A few number of control variables was included in the model because of little data availability. Other types of econometric models should be considered for robustness test and the study does not make a comparative analysis with other countries in the Eastern Europe. Therefore, future directions of research might refer to the use of panel data models at county level and a comparative analysis with other New EU Member States.

Acknowledgement

A part of this study related to data processing in MATLAB belongs to the research entitled “Econometric models in MATLAB”, which is part of the research plan of the Institute for Economic Forecasting of the Romanian Academy for 2024.

References

1. Anselin L (2005) Exploring spatial data with GeoDaTM: a workbook. *Center for spatially integrated social science* 165-223
2. Anvik C, Gjelstad K (2010) Just Google It!. BI Norwegian Business School
3. Askitas N, Zimmermann KF (2009a) Google econometrics and unemployment forecasting. *Applied Economics Quarterly* 55(2): 107-120. https://doi.org/10.2139/ssrn.1415585
4. Askitas N, Zimmermann KF (2009b) Googlemetrie und Arbeitsmarkt. *Wirtschaftsdienst* 89(7): 489-496. https://doi.org/10.1007/s10273-009-0957-0
5. Barreira N, Godinho P, Melo P (2013) Nowcasting unemployment rate and new car sales in south-western Europe with Google Trends. *NETNOMICS: Economic Research and Electronic Networking* 14(3):129-165. https://doi.org/10.1007/s11066-013-9082-8
6. Bea MD, Yi Y (2019) Leaving the financial nest: Connecting young adults’ financial independence to financial security. *Journal of Marriage and Family* 81(2): 397-414. https://doi.org/10.31235/osf.io/t6ar5
7. Borup D, Schütte ECM (2022) In search of a job: Forecasting employment growth using Google Trends. *Journal of Business & Economic Statistics* 40(1): 186-200. https://doi.org/10.2139/ssrn.3423124
8. Brave SA, Butters RA, Fogarty M (2020) Another Look at the Correlation Between Google Trends and Initial Unemployment Insurance Claims.
9. Buheji M (2019) Youth unemployment mitigation labs-an empathetic approach for complex socio-economic problem. *American Journal of Economics* 9(3): 93-105
10. Caperna G, Colagrossi M, Geraci A, Mazzarella G (2020) Googling unemployment during the pandemic: Inference and nowcast using search data. *JRC Working Papers in Economics and Finance* 2020/04:1-17 https://doi.org/10.2139/ssrn.3627754

11. Caperna G, Colagrossi M, Geraci A, Mazzarella G (2022) A babel of web-searches: Googling unemployment during the pandemic. *Labour Economics* 74:1-29 https://doi.org/10.1016/j.labeco.2021.102097
12. Chadwick MG, Sengül G (2015) Nowcasting the Unemployment Rate in Turkey: Let's Ask Google. *Central Bank Review* 15(3): 1-23.
13. Choi H, Varian H (2012) Predicting the present with Google Trends. *Economic Record 88*(s1): 2-9 https://doi.org/10.1111/j.1475-4932.2012.00809.x
14. D'Amuri F (2009) *Predicting unemployment in short samples with internet job search query data*. University Library of Munich, Germany.
15. Dilmaghani M (2019) Workopolis or The Pirate Bay: What Does Google Trends Say about the Unemployment Rate?. *Journal of Economic Studies* 46(2):422-445 https://doi.org/10.1108/jes-11-2017-0346
16. Doerr S, Gambacorta L (2020) *Identifying regions at risk with Google Trends: the impact of Covid-19 on US labour markets*. Bank for International Settlements 8.
17. Dumitrescu EI, Hurlin C (2012) Testing for Granger non-causality in heterogeneous panels. *Economic modelling* 29(4): 1450-1460 https://doi.org/10.1016/j.econmod.2012.02.014
18. Fajar M, Prasetyo OR, Nonalisa S, Wahyudi W (2020) Forecasting unemployment rate in the time of COVID-19 pandemic using Google trends data (case of Indonesia).
19. Fenga L, Son-Turan S (2020) Forecasting youth unemployment in the aftermath of the COVID-19 pandemic: the Italian case 1:1-28. https://doi.org/10.21203/rs.3.rs-74374/v1
20. Fondeur Y, Karamé F (2013) Can Google data help predict French youth unemployment?. *Economic Modelling* 30: 117-125 https://doi.org/10.1016/j.econmod.2012.07.017
21. Larson WD, Sinclair TM (2022) Nowcasting unemployment insurance claims in the time of COVID-19. *International Journal of Forecasting* 38(2): 635-647 https://doi.org/10.1016/j.ijforecast.2021.01.001
22. Lasso F, Snijders S (2016) The power of Google search data; an alternative approach to the measurement of unemployment in Brazil. *Student Undergraduate Research E-journal! 2*: 1-9.
23. McLaren N, Shanbhogue R (2011) Using internet search data as economic indicators. *Bank of England Quarterly Bulletin* 2: 134-140 https://doi.org/10.2139/ssrn.1865276
24. Oleksandr B (2010) *Can Google's search engine be used to forecast unemployment in Ukraine* (Doctoral dissertation, Kyiv School of Economics).
25. Pavlicek J, Kristoufek L (2015) Nowcasting unemployment rates with Google searches: Evidence from the Visegrad group countries. *PloS one* 10(5): 1-11https://doi.org/10.1371/journal.pone.0127084
26. Roopnarine KA, Spencer JD (2021) *Exploring the Use of Internet Searches to Predict Unemployment in Trinidad and Tobago*. Central Bank of Trinidad & Tobago.
27. Škuflić L, Vučković V (2018) The effect of emigration on unemployment rates: the case of EU emigrant countries. *Economic research-Ekonomska istraživanja* 31(1): 1826-1836 https://doi.org/10.1080/1331677x.2018.1516154
28. Sotis C (2021) How do Google searches for symptoms, news and unemployment interact during COVID-19? A Lotka–Volterra analysis of Google Trends data. *Quality & quantity* 55(6): 2001-2016 https://doi.org/10.1007/s11135-020-01089-0
29. Stewart K (2005) Dimensions of well-being in EU regions: Do GDP and unemployment tell us all we need to know?. *Social Indicators Research* 73(2): 221-246 https://doi.org/10.1007/s11205-005-2922-7

30. Su Z (2014) Chinese online unemployment-related searches and macroeconomic indicators. *Frontiers of Economics in China* 9(4): 573-605.
31. Suhoy T (2009) *Query indices and a 2008 downturn: Israeli data*. Bank of Israel.
32. Vicente MR, López-Menéndez AJ, 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
33. Yi D, Ning S, Chang CJ, Kou SC (2021) Forecasting unemployment using Internet search data via PRISM. *Journal of the American Statistical Association* 116(536): 1662-1673 https://doi.org/10.1080/01621459.2021.1883436
34. Yurevich MA, Akhmadeev DR (2021) Predicting the unemployment rate: Analyzing statistics on search engine query. *Terra Economicus* 19(3): 53-64 https://doi.org/10.18522/2073-6606-2021-19-3-53-64