Journal of Mathematics and Statistics Studies (JMSS)
www. al-kindipublisher.com/index.php/jmss

JMSS

The Trajectory of Corona-virus and Peoples' Concern in Africa: A Markov Switching Regression Model Based on Google Trends® Analysis

CHELLAI Fatih
Faculty of Economics, Commerce and Management; Ferhat Abbas University, Algeria
Corresponding Author: CHELLAI Fatih, E-mail: aith.chellai@univ-setif.dz

ARTICLE INFORMATION

Received: September 21, 2020
Accepted: October 02, 2020
Volume: 1
Issue: 1

KEYWORDS

Coronavirus, Google Trend, Markov Switching, Concern, Behavior

ABSTRACT

Coronavirus has greatly affected peoples' behavior, which led to an increase in fears and mental disorders in societies. The objective of the article is to estimate the effect of the virus spreading on the dynamic of people's concerns in four African countries (South Africa, Egypt, Nigeria, and Algeria). To reach such an objective, Markov switching (MS) models were estimated, where we used the Google Trends® and the new confirmed cases from March 13, 2020, to August 28, 2020. The findings revealed a weak positive correlation between Google Trends and the new confirmed cases in all countries. The MS estimation results showed a weak negative effect of the new confirmed cases on the waves of people's interest over the study period. We think that the effect of virus spreading is vanished by other factors (such as Media coverage). Nevertheless, the technical and methodological limitations, this study revealed the importance of web search tools (like Google Trends) in providing policy-makers utile data and information during periods of pandemic and health crisis to elaborate efficient awareness and prevention strategies.

1. Introduction

Since the first confirmed case in China in late 2019, Coronavirus disease is still the most up-to-date topic in the World, (Wang et al, 2020). This virus has nearly the same symptoms of SARS (*Severe Acute Respiratory Syndrome*) (Zhong et al, 2003), and MERS (*Middle East Respiratory Syndrome*); it can cause a simple cold as well as a serious respiratory infection such as pneumonia, causing fatal epidemics (Jin et al, 2020). The virus has rapidly spread to all seven continents at deferent levels. In the early time of the pandemic, the African countries had remained relatively unaffected comparing with other countries, (Nachega et al, 2020), unfortunately, since mid-March 2020 the situation has changed a lot, most of the countries have recorded a large spread of the virus in terms of the number of confirmed cases and the number of deaths.

From a different angle, daily statistics about the spread of the virus are a main source of disturbance, concern and stress for peoples over the World, see for instance: (Vinkers et al, 2020) and (Arslan et al, 2020). Scientists in different fields are working in an incessant way to study and analyze the patterns and effects of this virus, for example, and according to statistics provided by the Researchget platform, 95 180 of scientific papers (Articles, Preprints, Conference papers…) have been published from December 2019 to August 2020, see (Community Covid-19, 2020). Our study belongs to such initiatives for better investigation and analysis of this pandemic.

As a practical tool, we can measure the feeling, concern, and behavior of people in response to the spread of the virus by the Google Trends® platform; this tool, therefore, constitutes an interesting basis for carrying out observational epidemiological studies. Google Trends® could represent an interesting tool for providing healthcare professionals and authorities with useful information for the development of their healthcare structures. For example, the study of (Harorli & Harorli, 2014) identified for different regions of the world what were the main terms searched on the topic of oral pathologies over the past 10 years. Furthermore, different studies used the Google Trends® in other topics, e.g. (Cavazos-Rehg et al, 2015) applied the tool to

AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT
Your gateway to world-class research

monitor the use tendency of non-cigarette tobacco. (Schootman et al, 2015) examined the dynamics of cancer screening interests.

In the context of the coronavirus pandemic, an increasing number of articles and studies that used Google Trends® were published. (Higgins et al, 2020) revealed the existence of a strong positive potential of web search (in Google Trends and the Baidu Index) with coronavirus incidence rates. (Ana et al, 2020) aimed to examine the information-seeking responses to the virus spreading in the United States; they revealed that the level of the intention of people increased immediately after reporting results about the virus situation in the state, noted that the results of their study are limited at the first weeks of the epidemic. (Rovetta & Bhagavathula, 2020) focused on an infodemiological study by analyzed the behavior of people in web searching (mainly Google Trend and Instagram hashtags) related to the corona pandemic. They stated that the most affected countries had a higher rate of Google Trend searches queries. (Sharma & Sharma, 2020), analyzed the pattern of Google Trends web searching and the number of infected cases in eight countries, named (United States, Spain, Italy, France, United Kingdom, China, Iran, and India), the authors establish the existence of a positive correlation between the number of infected cases and the Google trend values have been recorded for eight major countries. We found also, (Ortiz-Martínez et al, 2020) tried to estimate and predict the coronavirus incidence using Google® trends. Noted also the study of (Sousa-Pinto et al, 2020), based on data from 17 countries, tried to check the effect of media coverage on the Google Trend, they conclude that the impact of virus spread is partially vanished by the Media-coverage.

Key limitations of these works are: *(i)* the relatively short study periods, which may hide a latent structural change of the relation between the spread of the virus (new infected cases) and the level of Google Trends index, *(ii)* Not using regression models to estimate the effect of the outbreak of the virus on people's concern. To the author's best knowledge, no study can be found that discuss such a topic in Africa. Inspired by the lines of these studies (cited above), we want to go forward in this topic, by estimating Markov Switching models to assess the dynamics of the potential effect of outbreak of Coronavirus on the trend of people's concerns in four African countries named: (South Africa, Egypt, Nigeria, and Algeria), these countries are ranked among the most affected by the virus in the continent. Mathematically, these models are based on a non-linear relationship, which implies that Google Trends indexes and new cases of Covid-19 in each country behave differently in each state (regimes). These models were developed in the financial and economical fields (Hamilton, 1989), nevertheless, several works of literature in public health and detection infectious diseases used them, see for instance, (Jung, 2006), (Lu et al, 2009) and recently (Amorós et al, 2020).

The rest of the paper is decomposed as follows: the second section presents a brief overview of the switching regression models based on Markov chains for the transition matrix among the different regimes estimated for the time series. The third section is devoted to exploratory analysis of the data used. The fourth section shows the results of the modeling, where a detailed discussion was given. The final section summarizes the results of this study and draws conclusions.

2. Methods

This section presents a brief overview of the Markov switching regression models. For more theoretical details, we recommend the following previous surveys about this statistical method: (Maddala, 1986), (Hamilton, 1994) and (Frühwirth-Schnatter, 2006). The use of MS in time series data is motivated by a situation where a dependent variable y_t may exhibits different states or regimes (r_t) over time, noted $k, i: 1, 2, \dots, K$. Switching modeling supposes that we have different regression models depending on k −regimes. The switching regimes mechanism mainly depends on: (i) specific assumptions for model construction and (ii) nature of *a priori* information (Full, incomplete, stochastic) (Uctum, 2007).

We assume a linear relationship between the dependent variable Y_t and explanatory variables X_t, Z_t, and we put,

$$Y_{kt} = \mu_{kt} + X'_{kt}\beta_k + Z'_t\theta + \varepsilon_t \qquad (1)$$

Where, μ_{kt}, X_{kt} and β_k: are means, switching regressors, and their relative coefficients (specific for each k-regime). Z_t and θ: are non-switching regressors and their corresponding coefficients. ε_t: are iid normally distributed. Further the specification in equation (1), in practice, when analyzing the Google Trends data of Covid-19, we don't know the switching mechanism among regimes. Mathematically, mixture-normal distributions modeling can be applied to, see for instance (Goldfeld & Quandt, 1973a), or the mechanism should be assumed to evolve from one regime to another according to an unknown transition probability, mainly under a Markov assumption, (**Engel & Hamilton, 1990**). We adopted the last option, and the full log-likelihood of the new model is given by,

$$l(\beta, \theta, \sigma, \boldsymbol{\psi}) = \sum_{t=1}^{T} log\left(\sum_{k=1}^{K} \sigma_k^{-1}\, \phi\left(\frac{Y_{kt} - X'_{kt}\beta_k + Z'_t\theta}{\sigma_k}\right)\right) . P(\boldsymbol{r_t} = \boldsymbol{k} \setminus \boldsymbol{\psi_{t-1}}) \qquad (2)$$

We assume that standard deviations are regimes dependents, ($\sigma_k \neq \sigma$). $\boldsymbol{\psi_{t-1}}$: resumes all information about regime transitions in the previous period ($t-1$). We define the regime probabilities (respecting the Markov first-order assumption) as,

$$P(r_t = j \setminus r_{t-1} = i) = p_{ij} \tag{3}$$

Which represents the probability transition from regime i to regime j, and verified:$\sum_{j=1}^{K} p_{ij} = 1$; this is the main feature of the stochastic matrix. Parameters that maximize the log-likelihood function (*equation 2*) are most often estimated by an application of the EM (Expectation-Maximization) algorithm, (Dempster et al, 1977). As another theoretical note in this section, (Psaradakis& Spagnolo, 2003) proposed a method to determine the number of regimes in switching regression models.

3. Data Selection and Description

3.1 Data selection

We describe briefly hereafter the strategy we followed in data selection, firstly for countries, then statistics of new confirmed cases of COVID-19 and finally the queries data from **Google Trends®.**

Country selection: we select four African countries ranked among the top ten most affected by the virus, up to August 30, 2020, the most affected country is South Africa by 622 551 as total cases and 13 981 total deaths, followed by Egypt (98 497 total confirmed cases), Nigeria and Algeria.

Covid-19 data: The statistics relatives of virus spreading are from the: worldometers.info website acceded on August 30, 2020, and the daily reports of the World Health Organization (WHO).

***Google* Trends®:** based on the studies of (Rovetta et al, 2020), (Higgins et al, 2020) and ((Sousa-Pinto et al, 2020), we retrieved the queries data via **Google Trends®** for these terms: (Corona, Coronavirus, Coronavirus, Covid-19, fever, Sars, Cough, Vomiting), which have been selected to estimate the trend of people's concerns in these countries over the study period. Further the English terminology and when we estimated for Algeria and Egypt, we selected the Arabic counterpart (using Google Translation) the following terms: ("السعال","الحمى","فيروس","كورونا"، "كوفيد19"، التقيؤ").

***Statistical programs*:** Analysis and estimation procedures were performed using **R** program (mainly: "MSwM", "forecast "and "ggplot2" packages) and **Eviews** (version 10.)

3.2 Descriptive analysis of the Data

Table 1 reports the summary statistics of the new cases and Google Trends time series of the four countries. As a measure of central tendency, we reported the median values, because, it is generally used to return the central tendency of skewed distributions. The median values for Google Trends trajectory are (38.79, 38.67, 28.36, and 24.93), respectively, for South Africa, Algeria, Nigeria, and Egypt; this means that half of the period from March 13, 2020, to August 28, 2020 (equivalently of 84 days), the indexes of Google Trends were below 50%. As a comparison between these countries, peoples in South Africa and Algeria have more tendencies in web searching for Covid-19 terminology compared with peoples in Nigeria and Egypt. The level of Google Trends dispersion is reported here by the Coefficient of Variation (CV); (for comparison between countries; because the means and medians are not equals). The same manner of interpretation can be made with the New Cases time series. The median daily new confirmed cases reported in South Africa is 1722 cases, followed by Egypt, Nigeria, and Algeria, respectively, with 391, 318, and 176 new cases; in other words, 84 days over the study period, the new cases in these countries have not exceeded the median values.

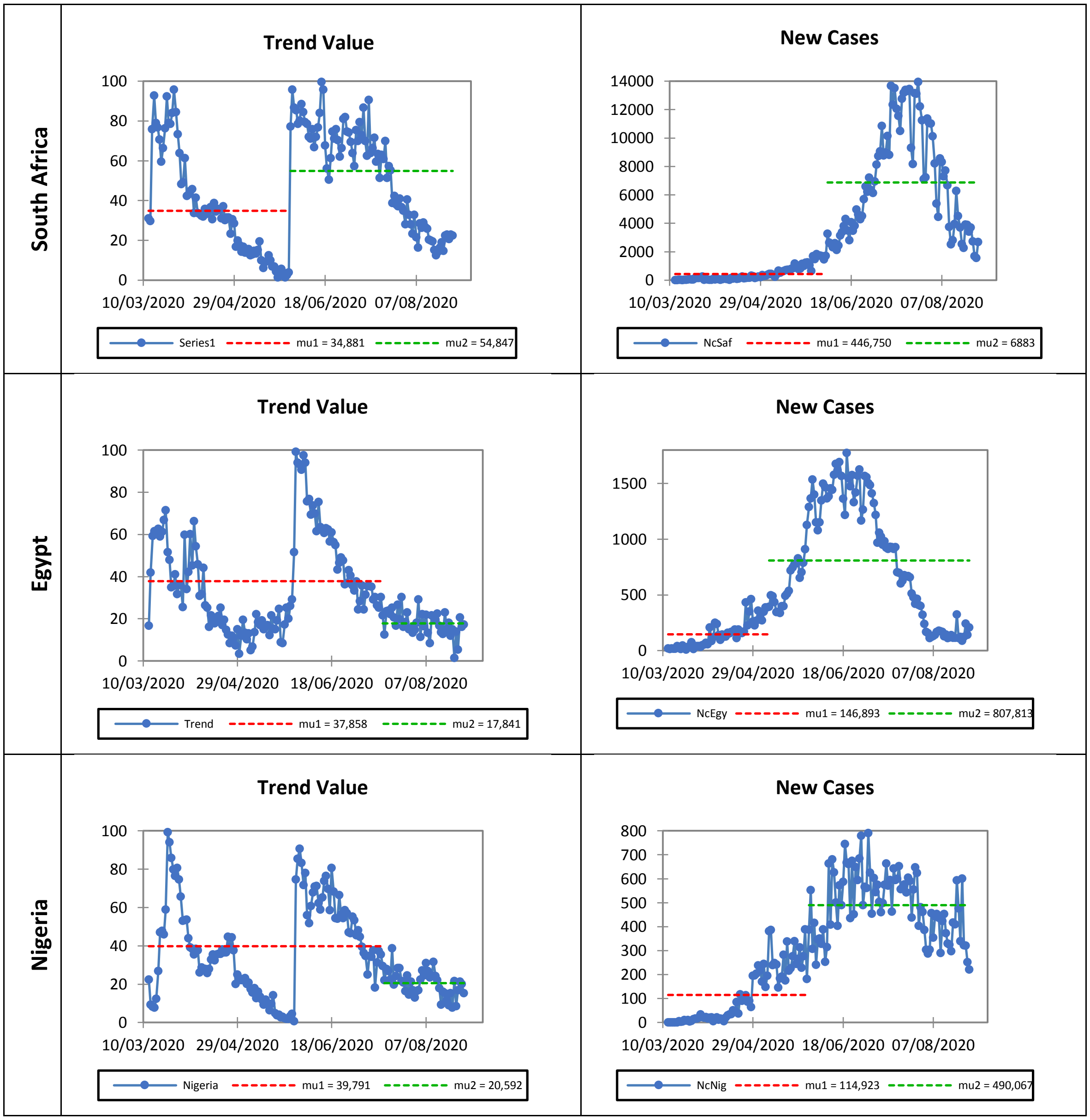
Table1: Summary Statistics of Google Trends and New cases time series

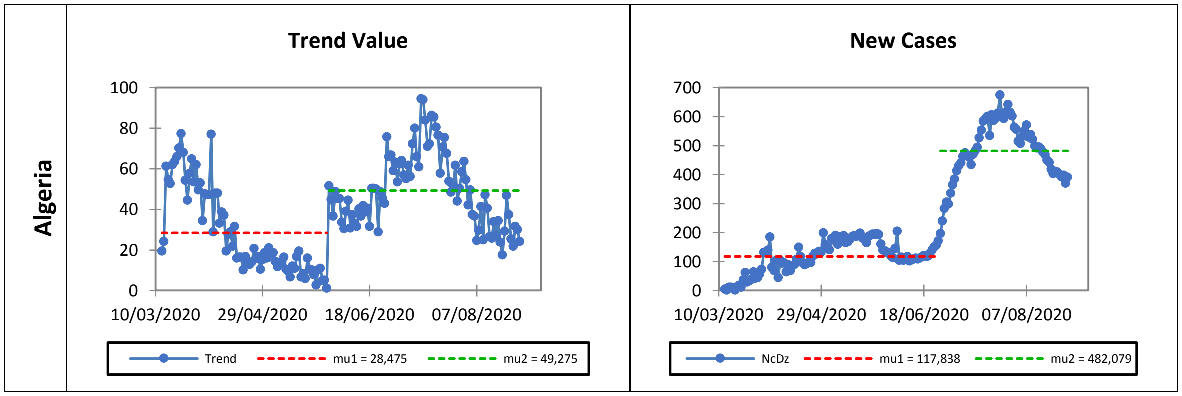
	Google Trends					**New Cases**				
	Median	CV	Skewness	Kurtosis	J.B	Median	CV	Skewness	Kurtosis	J.B
South Africa	38.79	1.68	0.15	1.72	12.10	1722	0.87	1.06	2.85	31.7
Egypt	24.93	1.49	1.12	3.56	37.25	391	1.10	0.72	2.08	20.4
Nigeria	28.36	1.52	0.72	2.72	15.21	318	1.39	0.08	1.80	10.3
Algeria	38.67	1.79	0.28	2.21	6.51	176	1.32	0.63	1.92	19.2

Source: https://www.worldometers.info/coronavirus/#countries ,for COVID-19 data. **Me**: Median, **SD:** Standard deviation, **CV**: Coefficient of variation, Sk: Skewness, **Ku**: Kurtosis, **JB**: Jarque-Bera test.

The Fisher skewness and kurtosis parameters were used to analyze the shape of data distribution; the skewness provides us information about the symmetric of the data compared with a normal distribution, whereas the kurtosis coefficient evaluates the dispersion of "extreme" values by reference to the normal distribution. Statistics in Table1 shows that the Google Trends data are asymmetric (skewed right) for Egypt and Nigeria and a relatively symmetric distributed in the case of South Africa and Algeria. Nearly the same data distribution is shown for the new-cases time series. In Figure1 we detect a sharp increase in the new confirmed cases from 15th march to mid-June 2020, and this for the four countries. After this sub-period, series of low trends in virus spreading could be observed especially in the case of South Africa and Egypt.

Figure.1 Evolution of Google Trends® and new cases of Covid-19 in: South Africa, Egypt, Nigeria, and Algeria over the study period.





Source: Plotted by Author using XLSTAT program.

Graphs in Figure1 show the dynamic of Google Trends queries of (Coronavirus terminology, see section 3.1) and the virus out-breaking which is measured by the new confirmed cases, in South Africa, Egypt, Nigeria, and Algeria. South Africa has recorded the peaked indexes on the Google wave of interest during the period from 27th March to 16th June 2020, for Algeria, the peaks have been registered on 22nd March and 13th July 2020.

The red and green lines traced referring to the change points detected in the eight-time series, where the Pettitt's test has been used (Pettitt, 1979). For all selected countries, two different regimes have been depicted for Google trends values and new cases of the virus. Through the estimation results of the Pettitt's test, the mean values for the first and second regimes of South Africa, Egypt, Nigeria, and Algeria are (34.88, 54.84), (37.83, 17.84), (39.79, 20.59), and (28.47, 49.27) respectively.

The y-axis of Google Trends plots reflects the proportion of web-searching for the different words selected in this study (see section 3.1 above); these proportions (or indexes) have been estimated in a specific region and period, compared to the region with the highest web-searching rates. For example, a value of 50 means that the keyword was used half as often in the affected region, and a value of 0 means that there is insufficient data for that keyword. Potential sources of variations in Google Trends among the four countries: connectivity index, volume of population, date of the first confirmed cases, and coverage in local (public & private) media…etc.

4. Estimation Results AND Discussion

4.1 Cross-correlation between Google Trends and Virus outbreak

Before starting regression modeling, firstly, the Cross-Correlation Functions (CCF) have been estimated and plotted in Figure2. The CCF function provides a set of sample correlations between tow time series x_{t+h} and y_t for different lags $h = 0, \pm1, \pm2, \pm3$, it's a generalization of the simple correlation coefficient at level (*i.e.* when $h = 0$, r_{xy}). As a result showed in Figure2, the max correlation coefficients are surrounding the lags ($h = -20$ to $h = -7$) in the case of South Africa, indicating that an increasing (in average) of new cases of Coronavirus is very likely to lead to a decrease in Google Trends of people in this country about ten days later and vice-versa. For Egypt, we found a symmetric relationship between Google Trends (as output) and new cases of Covid-19 (as a driver), the peaks of correlations are visible at lags ($h = 0, \pm1, \pm2$). We found poor (weak) correlations between selected terminology of search in Google Trends and the incidence of Coronavirus in Nigeria.

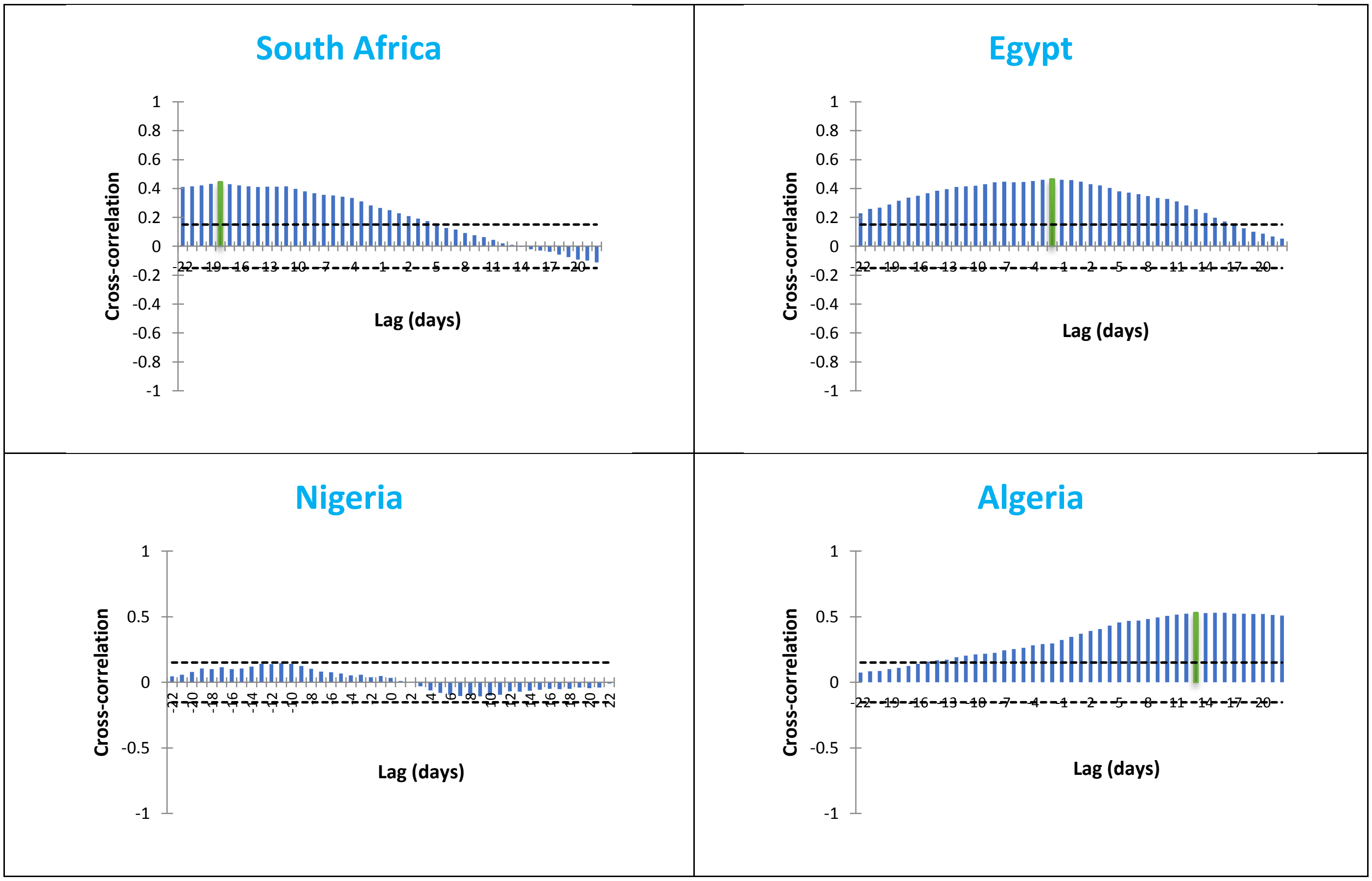


Figure2 Cross-Correlation of Google Trends of Covid-19 (terminology detailed in the above section) with New cases of Coronavirus.

We examined the CCF functions to assess that the spread of virus could be a significant leading factor of the Google Trends indexes; after this step the Broock- Dechert- Scheinkman (BDS) test have been employed, (Broock et al, 1996), this test is based on the null hypothesis that the time series are independent and identically distributed (*iid*). According to the calculated test-statistics for the four relationships (Google Trends **-** New cases) the null hypothesis of independence has been rejected, which implies that the data are *iid*, and the relationship between Google Trends and New cases is not-linear for all countries. Such results made the Markov Switching Regression models more suitable (*or at least a competitive approach*) to estimate and analyze the relationship between the Google Trends and the new cases of COVID-19.

The estimation results of the Markov models are summarized in Table 2. *First*ly, noted, we confirmed the assumption set in *Equation* (**2**) about the regimes' dependence of the standard deviations; and we fitted two-States Markov switching models, in mean and in variance. *Second* note, for the "*Intercept*" estimated coefficients (which are the *regime specific means*), the results indicate that for South Africa, Egypt, and Algeria the highest Google Trends indexes correspond to *Regime1*, whereas in the case of Nigeria the highest indexes were in *Regime2*; in contrast, the lowest Google Trends of (South Africa, Egypt and Algeria) are in *Regime2* and (Nigeria) in *Regime1.*

Table2. Results of Markov Switching regressions estimation

Country	Regimes	Variable	Coef (SE)	*P. Value*	*σ*	R^2	AIC	BIC	Log-Like
South Africa	(1)	New Cases	-0.0007(0.0001)	***	8.63	0.119	7.73	7.86	-621.31
		Mean	75.12 (1.53)	***					
	(2)	New Cases	0.0007 (0.0003)	**	8.93	0.071			
		Mean	31.95 (1.17)	***					
Egypt	(1)	New Cases	0.012 (0.002)	***	10.93	0.358	7.22	7.43	-551.45

		Intercept	53.27 (2.11)	***					
	(2)	New Cases	0.011 (0.001)	****	4.51	0.561			
		Intercept	28.69 (0.72)	***					
Nigeria	(1)	New Cases	-0.007 (0.005)	-	10.16	0.021	7.02	7.13	-649.97
		Intercept	31.82 (2.12)	***					
	(2)	New Cases	-0.02 (0.009)	*	11.77	0.123			
		Intercept	74.09 (5.01)	***					
Algeria	(1)	New Cases	0.049 (0.007)	***	6.75	0.382	7.62	7.75	-623.98
		Intercept	49.68 (1.84)	***					
	(2)	New Cases	0.064 (0.004)	***	11.71	0.742			
		Intercept	15.31 (1.571)	***					

Notes: ***, ** and * indicate (respectively) significance at 1%, 5% and 15%. The values in (.) are standard error.

Models Comparisons and Adequacy

For model adequacy, we use the multiple R squared, see Table3, the small values of R^2 indicate a weak contribution of the COVID-19 spreading in explaining the variation of the Google Trends during the analyzed period. Results in Table3 indicate an insignificant effect of the spread of virus on Google Trends in South Africa and Nigeria; the exception was for Algeria and Egypt modeling, precisely in the *Regime2*, where we estimated that 56.1% and 74.2% of the variation in Google Trends is explained by the dynamics of the new cases. Noted that the poorest of Markov switching model fitting in the case of Nigeria confirms –partially- the weak cross correlations estimation plotted in Figure. Noted that, we estimated four models for each country: (1) Models without shift neither in mean and variance (2) Models with shift in means (3) Models with shift in variance (4) Models with a shift in mean and variance. The last models were the optimal ones for all countries, where the log-likelihood functions, Akaike (**AIC)** and Bayesian (**BIC**) information criteria were used for comparison. Furthermore, the residual analysis implied that they follow a normal distribution (according to Jarque-Bera statistics), and are homoscedastic (using ARCH tests).

In terms of regressors coefficients (in Table2), the statistically significant effect of the new confirmed cases on the Google Trend values has been estimated for all countries and over the two regimes. Nevertheless, the values of the coefficients of the New cases' variable confirmed that the effect of the virus spreading on the Web search was very weak for all countries except for Algeria. In light of such findings, we think that other hidden and not adjusted factors have (direct and/or indirect) effects on the Google Trends variations in these countries.

For the case of South Africa and Nigeria, the two regimes recorded nearly the same pattern of volatility in term of the Google Trends indexes (see the column of $\boldsymbol{\sigma}$ in Table2). Contrary, for Egypt and Algeria, the two regimes showed a distinctive level of volatility, for example, in Algeria *Regime1* has low volatility ($\boldsymbol{\sigma_1 = 6.75}$), while *Regime2* is having high volatility ($\boldsymbol{\sigma_2 = 11.71}$); this implies that for the Algerian case peoples present homogenous pattern of Google Trend distribution during periods of high web search (in Regime1) and heterogeneous behavior during periods of weak web search (in Regime2). We see that the inverted pattern can be reported for Egypt.

Table3. Transition Matrix and expected durations of staying in the two-regimes

Country	Transition matrices (*)	Expected Duration (**)	
		Regime (1)	Regime (2)
South Africa	$\begin{array}{cc} & \begin{array}{cc} r_1 & r_2 \end{array} \\ \begin{array}{c} r_1 \\ r_2 \end{array} & \begin{pmatrix} 0.968 & 0.032 \\ 0.018 & 0.982 \end{pmatrix} \end{array}$	**31.81**	**56.99**
Egypt	$\begin{array}{cc} & \begin{array}{cc} r_1 & r_2 \end{array} \\ \begin{array}{c} r_1 \\ r_2 \end{array} & \begin{pmatrix} 0.978 & 0.022 \\ 0.048 & 0.952 \end{pmatrix} \end{array}$	**69.71**	**21.03**
Nigeria	$\begin{array}{cc} & \begin{array}{cc} r_1 & r_2 \end{array} \\ \begin{array}{c} r_1 \\ r_2 \end{array} & \begin{pmatrix} 0.953 & 0.047 \\ 0.014 & 0.986 \end{pmatrix} \end{array}$	**68.64**	**21.28**
Algeria	$\begin{array}{cc} & \begin{array}{cc} r_1 & r_2 \end{array} \\ \begin{array}{c} r_1 \\ r_2 \end{array} & \begin{pmatrix} 0.976 & 0.014 \\ 0.019 & 0.981 \end{pmatrix} \end{array}$	**53.48**	**52.96**

Notes: **(**)** based on constant Markov transition probabilities**.** The expected duration is in ***(days).***

Table3 summarizes the estimation of the transition probability matrix and the expected durations for the four countries. From the transition matrices, we see that there is a weak dependency between *Regime1* and *Regime2* for all countries. Such dependency is reported by the transition probabilities: p_{12} and p_{21} (in blue in the transition matrices). In contrast, there is a higher probability of staying (or persistence) in the origin regime, which are p_{11} and p_{22}. The corresponding transition probabilities are 0.968, 0.978, 0.953, and 0.976 for Regime1, respectively, in: South Africa, Egypt, Nigeria, and Algeria, and approximately the same probabilities for Regime2 in all countries. As a main result in this point, we found that for all countries except Egypt, Regime2 is more persistent than Regime1. The estimation of the expected durations to remain in regime1 are (31.81, 69.71, 68.64, and 53.58 days) and (56.99, 21.03, 21.28, and 52.96 days) in regime2, respectively, in South Africa, Egypt, Nigeria and Algeria.

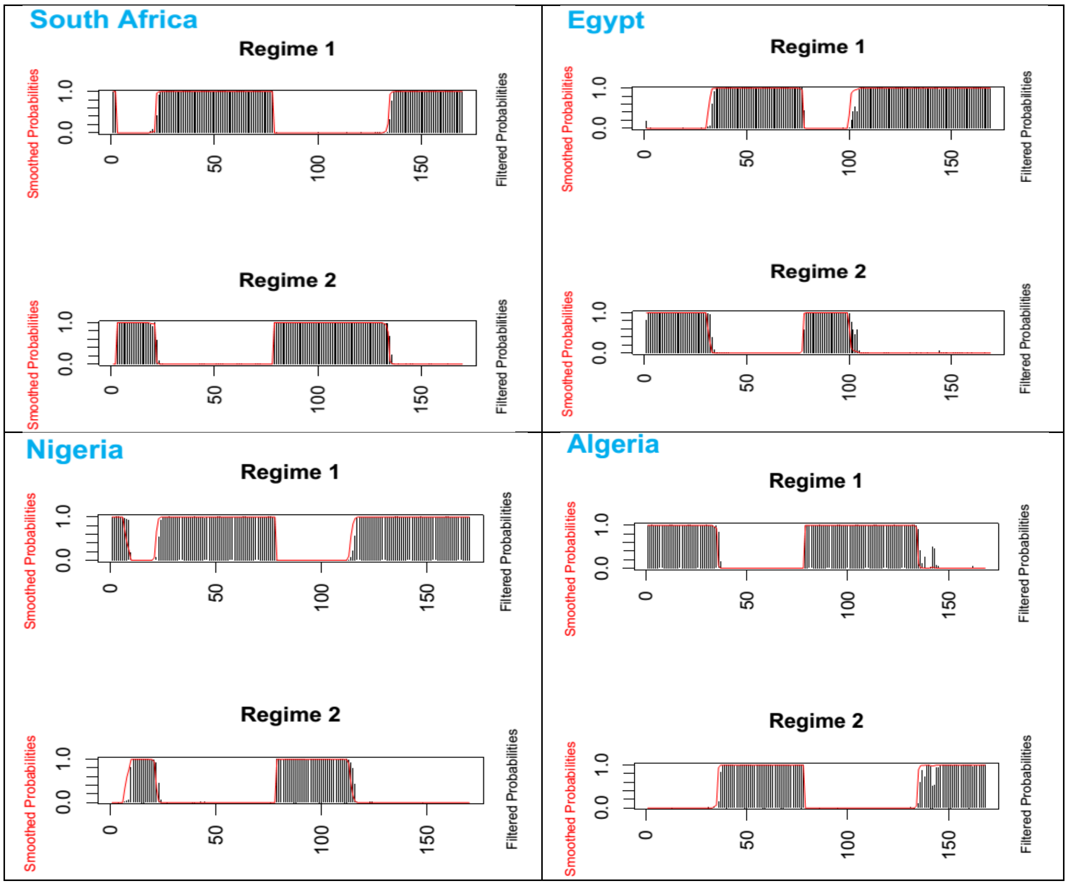


Figure3. Plots of regimes probabilities, *in red:* is the smoothed probability, *in black*: the filtered probabilities; $1 - p(r_t = i)$, with $i: 1,2.$ for each regime.

Plots in Figure3 are a visual representation of the dynamics of dependence between the two regimes, the smoothed probabilities showed in figure are the best tool to investigate and determine the breaks dates and regimes stability in Google Trends data. Moreover, these plots depict the filtered probabilities and smoothed probabilities for the two regimes and the four countries. The separated within-regimes indicates the dependence weakness between them (see p_{11} and p_{22} in Table1). We conclude that the Google Trends about COVID-19 in Egypt and Nigeria are more stable (in terms of duration and volatility) in Regime1 compared to Regime2.

A possible explication for this switching behavior, illustrated thorough Regime2 in Figure3 is the spreading of the economical and social effects of this virus, a fact that push peoples to get worried about the daily problems like: employments, schools, travels. Moreover, switching regimes from the highest to the lower indexes of Google Trends over the study period implies that awareness and concern of peoples in these countries follow a chaos pattern in parallel with to the evolution of the virus. This result may confirm partially the existence of others factors, especially media coverage, that affect the Google Trends indexes rather than the statistics of the virus, this idea has been well established by (Sousa-Pinto et al., 2020). This fact is well recognized when we analyze the data of Google queries about virus terminology before dates when first confirmed cases were declared in these countries. The first confirmed case was at (March 6th, February 15th, February 28th, and February 25th,

2020) respectively for: (South Africa, Egypt, Nigeria, and Algeria). The biggest mean values of Google Trend estimated before these dates imply that the data generating process of Google Trend is affected by other factors than the virus spreading.

As final notes in this study, the big challenge with such a study reside is the reliability of the data generated by Google Trends which mentioned by (Cervellin *et al*, 2017), precisely in epidemiology domains; because it gives a relative and not absolute search volume, furthermore, this tool only relates to a population with internet access and using Google® as a search engine; and for this point, we note the utility to use trends data from for example Facebook, Twitter, etc. Another challenge (and limitation) to use the data provided by Google Trends® reside in confusion to determine precisely the population characteristics targeted by the realizing study. The study needs more extension by including other factors to explain better the dynamics in Web searching relative to the Coronavirus pandemic.

5. Conclusion

This study deals with the estimation and analysis of the relationship between Google Trends and the spread of Coronavirus in four African countries (South Africa, Egypt, Nigeria, and Algeria). To investigate this relationship, at the first step, cross-correlation functions were estimated and plotted, in second step, two-state Markov Switching in intercept and in variance models were fitted. Summing up the empirical results, we found a positive correlation between the two variables. This study extended in some points the work of (Higgins et al, 2020) for some African countries, where our findings were largely in concordance with. Summing up the results, it can be concluded that the spread of the virus has a weak effect on Google Trends indexes for all countries except Nigeria. As policy implications, the decision-makers in the health systems and public media should elaborate psychological and awareness strategies to reduce the psychological and mental effects of this pandemic.

Financial support: This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Conflict of interest: None.

References

[1] Alvarez, R., Camacho, M., & Ruiz, M. (2019). Inference on filtered and smoothed probabilities in Markov-switching autoregressive models. *Journal of Business & Economic Statistics*, *37*(3), 484-495.DOI: 10.1080/07350015.2017.1380032

[2] Amorós, R., Conesa, D., López-Quílez, A., & Martinez-Beneito, M. A. (2020). A spatio-temporal hierarchical Markov switching model for the early detection of influenza outbreaks. *Stochastic Environmental Research and Risk Assessment*, 1-18.

[3] Ana I. Bento, Thuy Nguyen, Coady Wing, Felipe Lozano-Rojas, Yong-Yeol Ahn, Kosali Simon.Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases.Proceedings of the National Academy of Sciences May 2020, 117 (21) 11220 11222; DOI: 10.1073/pnas.2005335117.

[4] Arslan, G., Yıldırım, M., Tanhan, A., Buluş, M., & Allen, K. A. (2020). Coronavirus stress, optimism-pessimism, psychological inflexibility, and psychological health: Psychometric properties of the Coronavirus Stress Measure. *International Journal of Mental Health and Addiction*, 1.

[5] Broock, W. A., Scheinkman, J. A., Dechert, W. D., & LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric reviews*, *15*(3), 197-235.

[6] Cavazos-Rehg PA, Krauss MJ, Spitznagel EL, Lowery A, Grucza RA, Chaloupka FJ, (2015). Monitoring of non-cigarette tobacco use using Google Trends. Tob Control. 24(3):249-55. 61.

[7] Cervellin, G., Comelli, I., & Lippi, G. (2017). *Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. Journal of Epidemiology and Global Health, 7(3), 185–189.* doi:10.1016/j.jegh.2017.06.001.

[8] Community Covid-19, (2020). Retrieved September 5, 2020, from Researchgate Platform, https://www.researchgate.net/community/COVID-19.

[9] Engel, C. et J. D. Hamilton (1990), « Long Swings in the Dollar : Are They in the Data and Do Markets Know It? », *American Economic Review*, 80(4) : 689- 713.

[10] Goldfeld, S. M. et R. e. Quandt (1973a), « A Markov Model for Switching Regressions », *Journal of Econometric*, 1 : 3-16.

[11] Hamilton J.D. (1989). A New Approach to the Economic Analysis of Nonstionary Time Series and the Business Cycle. Econometrica 57: 357-384.

[12] Hamilton, J.D. (1994). Time Series Analysis. Princeton University Press.

[13] Harorli OT, Harorli H. Evaluation of internet search trends of some common oral problems, 2004 to 2014. Community Dent Health. sept 2014;31(3):188-92.

[14] Jin, X., Lian, J. S., Hu, J. H., Gao, J., Zheng, L., Zhang, Y. M., ... & Yu, G. D. (2020). Epidemiological, clinical and virological characteristics of 74 cases of coronavirus-infected disease 2019 (COVID-19) with gastrointestinal symptoms. *Gut*, *69*(6), 1002-1009.

[15] Jung, J. (2006). Estimating Markov transition probabilities between health states in the HRS dataset. *Indiana University*, 1-42.

[16] Lee, S. A. (2020). Coronavirus Anxiety Scale: A brief mental health screener for COVID-19 related anxiety. *Death studies*, *44*(7), 393-401.

[17] Lu, H. M., Zeng, D., & Chen, H. (2009). Prospective infectious disease outbreak detection using Markov switching models. *IEEE Transactions on Knowledge and Data Engineering*, *22*(4), 565-577.
[18] Nachega, J., Seydi, M., & Zumla, A. (2020). The Late Arrival of Coronavirus Disease 2019 (COVID-19) in Africa: Mitigating Pan-continental Spread. *Clinical Infectious Diseases*, *71*(15), 875-878.
[19] Ortiz-Martínez, Y., Garcia-Robled, J. E., Vásquez-Castañeda, D. L., Bonilla-Aldana, D. K., & Rodriguez-Morales, A. J. (2020). Can Google® trends predict COVID-19 incidence and help preparedness? The situation in Colombia. *Travel Medicine and Infectious Disease*.
[20] Pettitt A.N. (1979). A non-parametric approach to the change-point problem. *Appl. Statist.*, 28(2), 126-135.
[21] Psaradakis, Z. and Spagnolo, N. (2003), ON THE DETERMINATION OF THE NUMBER OF REGIMES IN MARKOV-SWITCHING AUTOREGRESSIVE MODELS. Journal of Time Series Analysis, 24: 237-252. doi:10.1111/1467-9892.00305.
[22] Rovetta A, Bhagavathula AS. (2020). Global Infodemiology of COVID-19: Analysis of Google Web Searches and Instagram Hashtags J Med Internet Res;22(8):e20673 DOI: 10.2196/20673 PMID: 32748790.
[23] Schootman M, Toor A, Cavazos-Rehg P, Jeffe DB, McQueen A, Eberth J, (2015). The utility of Google Trends data to examine interest in cancer screening. BMJ Open;5(6):e006678.
[24] Sharma M, Sharma S. (2020).The Rising Number of COVID-19 Cases Reflecting Growing Search Trend and Concern of People: A Google Trend Analysis of Eight Major Countries. *J Med Syst*;44(7):117. Published 2020 May 20. doi:10.1007/s10916-020-01588-5.
[25] Sousa-Pinto B, Anto A, Czarlewski W, Anto JM, Fonseca JA, Bousquet J ,(2020). Assessment of the Impact of Media Coverage on COVID-19–Related Google Trends Data: Infodemiology Study J Med Internet Res;22(8):e19611 URL: https://www.jmir.org/2020/8/e19611 doi: 10.2196/19611
[26] Uctum, R. (2007). Économétrie des modèles à changement de régimes : un essai de synthèse. *L'Actualité économique*, *83* (4), 447–482. https://doi.org/10.7202/019389ar.
[27] Vinkers, C. H., van Amelsvoort, T., Bisson, J. I., Branchi, I., Cryan, J. F., Domschke, K., ... & van der Wee, N. (2020). Stress resilience during the coronavirus pandemic. *European Neuropsychopharmacology*.
[28] Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *The Lancet*, *395*(10223), 470-473.
[29] Zhong, N. S., ZHENG, B. J., LI, Y. M., *et al. (2003).* Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, 374 People's Republic of China, in February. *The Lancet*, 2003, vol. 362, no 9393, p. 1353-1358.