CARMA 2024

6th Int. Conf. on Advanced Research Methods and Analytics
Universitat Politècnica de València, València, 2024
DOI: https://doi.org/10.4995/CARMA2024.2024.17503

Nowcasting food insecurity interest Google Trends data

Nicola Caravaggio[1], Bia Carneiro[2], Giuliano Resce[1]

[1]Department of Economics, University of Molise, Campobasso, Italy, [2]Bioversity International, Rome, Italy.

How to cite: Caravaggio, N.; Carneiro, B. ; Resce, G. 2024. Nowcasting food insecurity interest Google Trends data. In: 6th International Conference on Advanced Research Methods and Analytics (CARMA 2024). Valencia, 26-28 June 2024. https://doi.org/10.4995/CARMA2024.2024.17503

Abstract

This research explores the potential of Google Trends (GT) data as a tool for generating a daily index of food insecurity at the national level, focusing on regions monitored by the Famine Early Warning Systems Network (FEWS NET) and the Global Fragility Act (GFA). Drawing inspiration from previous studies on GT's predictive capabilities, the authors employ Natural Language Processing (NLP) to analyse food security reporting from FEWS NET documents. We identify key predictors of food insecurity using a LASSO regression approach and construct a daily economic sentiment index (DESI) for each country. Unlike traditional methods, the study considers multiple languages and weights search terms based on LASSO coefficients. The resulting Synthetic Search Interest (SSI) index for food insecurity demonstrates a statistically significant correlation with FAO's share of the population in severe food insecurity, affirming GT's potential as a monitoring tool. The research contributes a novel methodology and insights into leveraging real-time data for early warnings in food security.

Keywords: *food insecurity, Google trends, early warnings, Natural Language Processing*

1. Introduction

Official Statistics indicators regarding food, nutrition, and livelihood security outcomes are typically only available with a reporting lag of several weeks and are often revised a few months later. Nowadays, several sources of data on real-time economic activity are available from private sector companies such as Alphabet. An example is Google Trends (GT), a real-time daily, weekly, and monthly index of the volume of user queries on Alphabet's search engine Google. Such user search patterns are often correlated with various socio-economic indicators and may be helpful for short-term prediction and nowcasting.

This research aims to provide a strong tool able to generate a daily index of food insecurity at national level by relying on GT data. Considering that threats to food security affect mostly fragile and food insecure regions, the presented analysis has been developed on countries monitored by the Famine Early Warning Systems Network (FEWS NET) and the Global Fragility Act (GFA). Henceforth, GT may present an important tool for early warnings of food insecurity in highly stressed areas. Ginsberg et al. (2009) pioneered the use of GT in research studies. Their groundbreaking work showcased how GT could effectively monitor and forecast the progression of influenza ahead of the official reports from the Centers for Disease Control and Prevention (CDC) in the United States. Other authors have stressed the predictive potentiality of GT, especially regarding the present: "[w]e are not claiming that Google Trends data help predict the future. Rather we are claiming that Google Trends may help in predicting the present" (Choi & Varian, 2009) [p. 2].

2. The potentialities in the use of GT

Researchers are increasingly relying on GT under manifold circumstances (Jun, Yoo, & Choi, 2018). Choi and Varian (2009) have shown how GT data can help predict initial claims for unemployment benefits in the United States. Furthermore, Askitas and Zimmermann (2009) and Suhoy (2009) performed similar analyses stressing the potential of GT in Germany and Israel, respectively. Conversely, Nagao, Takeda, and Tanaka (2019) stressed some limitations in using GT to nowcast unemployment in the US. By working on retail, automotive, and home sales topics, Choi and Varian (2012) also demonstrated how seasonal autoregressive (AR) models and fixed-effects models that includes relevant GT variables tend to outperform models that exclude these predictors. Several studies used GT data within the financial market to predict, for example, the direction of stock market through neural networks trained with GT data (Fan, Chen, & Liao, 2021; Hu, Tang, Zhang, & Wang, 2018) or the identification of "early warnings signs" in financial markets (Petropoulos, Siakoulis, Stavroulakis, Lazaris, & Vlachogiannakis, 2022; Preis, Moat, & Stanley, 2013). Extensive application of GT data is also found within the field of epidemiology, for example, as source of real-time influenza surveillance (Broniatowski, Paul, & Dredze, 2013). Moreover, GT data is not confined solely to predictive areas but also as a support for user geolocation of Twitter (now X) data (Zola, Ragno, & Cortez, 2020). Nonetheless, these approaches and the general use of GT are not free from critiques (Cook, Conrad, Fowlkes, & Mohebbi, 2011; Lazer, Kennedy, King, & Vespignani, 2014).

During the outbreak of the Covid 19 pandemic, several studies investigated the phenomenon through the lens of GT (Kornellia & Syakurah, 2023). Kurian et al. (2020) evidenced the high correlation between Covid cases among US states and 10 keywords searched on GT; Liu et al. (2022) through a prophet model showed how GT of Covid related terms represented important predictors in investigating the number of cases among US states; Lampos et al. (2021) showed how online searches precede the number of confirmed cases and deaths by 16.7 and 22.1 days, respectively; Brunori and Resce (2020) relied on GT data to estimate a prediction model for the Italian case. Those studies generally demonstrate how online search data can be used to develop complementary public health surveillance methods in conjunction (not substitution) with more established approaches.

3. Data and methodology

Our work starts from a Natural Language Processing (NLP) analysis of food security reporting based on 1,414 publicly available documents from FEWS NET, covering 33 countries. The development of a custom taxonomy of food insecurity identified three groups of topics, namely (i) hazard/shocks, (ii) food security indicators, and (iii) food, nutrition, and livelihood security outcomes, each in turn composed of different sub-topics. A topic-matrix was created for the occurrence of each sub-topic among the list of country-year reports. For the purposes of this study, we rely only on the first topic-group (hazards/shocks) as it considers terms which effectively represent predictors of food insecurity. For example, it makes a more lot more sense in case of a food crisis that people would search on Google for topics which may effectively be identified as shocks (*e.g.*, conflict, exchange rate) rather than more general, outcome-related terms associated to this phenomenon (*e.g.*, food production, hunger). Topics within the hazards/shocks theme have been divided into five groups: climate, conflict, markets, diseases, and governance. We considered a total number of 47 countries, mostly located in Sub-Saharan Africa over a 10-year period spanning from January 1, 2013, to December 31, 2023.

In the first step of our analysis, we identified the most important predictors of food insecurity through a data driven approach using a LASSO regression on the classification results from the NLP model. Then, we selected the top ten positive features after excluding those disease-related, such as Covid-19 or Ebola, due to their time or geographic specificity. The necessity to identify only the most important features is twofold: (i) through the LASSO we were able to determine a feature's importance with relative coefficients that we used eventually to weight our GT search terms; (ii) the necessity to automate the download procedure from GT lead to a necessary balancing between data granularity and feature selection[1].

After determining the core predictors of food insecurity represented in FEWS NET reporting, we followed the daily economic sentiment index (DESI) approach proposed by Eichenauer, Indergand, Martinez, and Sax (2022). For each country, a long-run frequency-consistent daily trend of food insecurity was constructed. However, our approach is different in two major aspects. First, rather than considering only one specific language, we preferred to retrieve word search data by considering all national languages of our sample of countries, in addition to the most widely spoken languages in the world not included among the national languages of the countries analysed. This means that in each country, the search for a specific word has been replicated for 25 different languages. Results were then aggregated in order to constitute, for

[1] The download procedure has been carried out with R using the package gtrendsR.

each topic, a country-specific index. Second, to aggregate search terms into a single indicator, we preferred to weight each topic by the corresponding scaled coefficients obtained from the LASSO and then average results, rather than use a principal component analysis (PCA). In fact, we noticed that trying to create a synthetic index starting from trends of numerous words (up to 10), the PCA may not represent the most suitable tool since it often aggregates into an index of disinterest, hence assuming a pattern generally opposite to what is shown by their components.

4. Final outcome

The final time series constructed for each country represents a synthetic search interest (SSI) index for food insecurity based on GT data. In the final step of our procedure, we applied some further weights allowing for cross-country comparison. The first weight has been constructed by considering the worldwide GT interest of each topic while the second one relied on yearly data of internet penetration (WB, 2024) (inverse) to compensate for the digital divide among countries.

Lastly, we validate our data by performing a comparison between our SSI index and the share of population in severe food insecurity provided by FAO (2024). Results show a statistically significant and positive correlation between the two data suggesting GT may effectively help in representing a tool for monitoring food insecurity trends worldwide. The whole procedure adopted is graphically synthesized in *Figure **1*** while an example of SSI for Tanzania is shown in *Figure **2***.

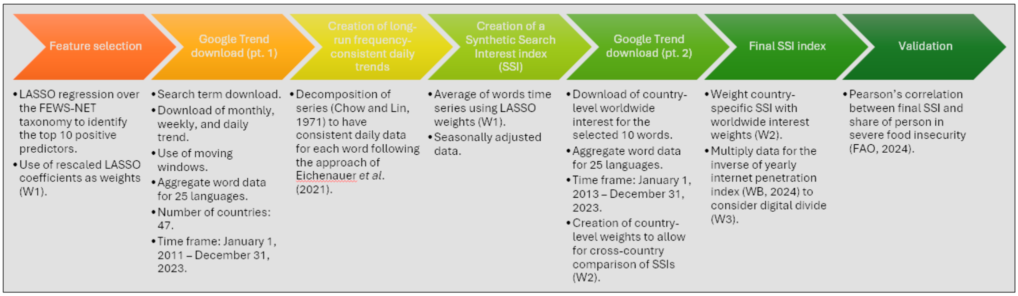

Figure 1. Cross-country synthetic search interest (SSI) creation procedure.

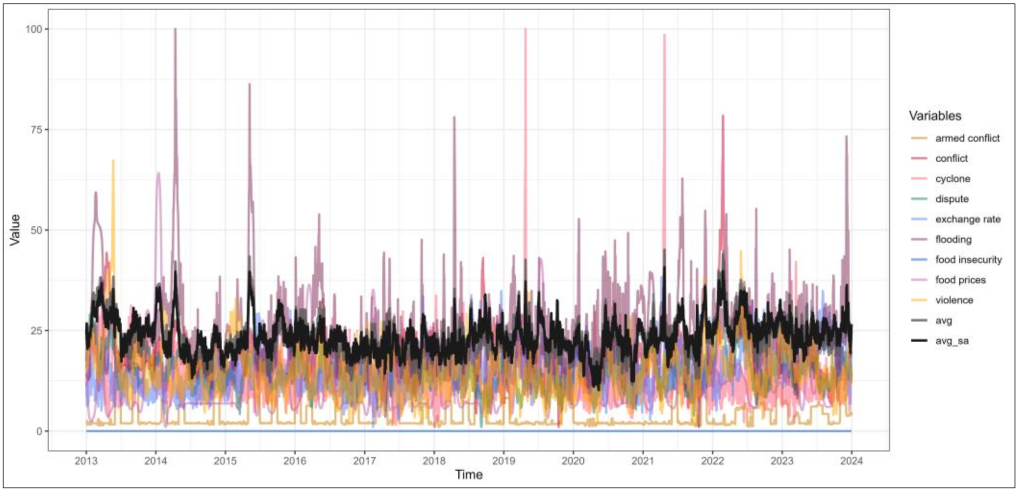

Figure 2. Example of synthetic search interest (SSI) for Tanzania.

References

Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. Applied Economics Quarterly, 55(2), 107–120.

Broniatowski, D. A., Paul, M. J., & Dredze, M. (2013). National and local influenza surveillance through twitter: an analysis of the 2012-2013 influenza epidemic. PloS one, 8(12), e83672.

Brunori, P., & Resce, G. (2020). Searching for the peak google trends and the covid-19 outbreak in italy. (SERIES working papers, No. 04/2020)

Choi, H., & Varian, H. (2009). Predicting initial claims for unemployment benefits. Google Inc, 1 (2009), 1–5.

Choi, H., & Varian, H. (2012). Predicting the present with google trends. Economic record, 88, 2–9.

Cook, S., Conrad, C., Fowlkes, A. L., & Mohebbi, M. H. (2011). Assessing google flu trends performance in the United States during the 2009 influenza virus a (h1n1) pandemic. PloS one, 6(8), e23610.

Eichenauer, V. Z., Indergand, R., Mart´ınez, I. Z., & Sax, C. (2022). Obtaining consistent time series from google trends. Economic Inquiry, 60(2), 694–705.

Fan, M.-H., Chen, M.-Y., & Liao, E.-C. (2021). A deep learning approach for financial market prediction: Utilization of google trends and keywords. Granular Computing, 6, 207–216.

FAO. (2024). FAOSTAT. Rome, IT. (Food and Agriculture Organization of the United Nations.

Retrieved from: http://www.fao.org/faostat/en/#data [Accessed February 1, 2024])

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. Nature, 457(7232), 1012–1014.

Hu, H., Tang, L., Zhang, S., & Wang, H. (2018). Predicting the direction of stock markets using optimized neural networks with google trends. Neurocomputing, 285, 188–195.

Jun, S.-P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using google trends: From the perspective of big data utilizations and applications. Technological forecasting and social change, 130, 69–87.

Kornellia, E., & Syakurah, R. A. (2023). Use of google trends database during the covid-19 pandemic: systematic review. Multidisciplinary Reviews, 6(2), 2023017–2023017.

Kurian, S. J., Alvi, M. A., Ting, H. H., Storlie, C., Wilson, P. M., Shah, N. D., . . . others (2020).

Correlations between covid-19 cases and google trends data in the United States: A state-by-state analysis. In Mayo clinic proceedings (Vol. 95, pp. 2370–2381).

Lampos, V., Majumder, M. S., Yom-Tov, E., Edelstein, M., Moura, S., Hamada, Y., . . . Cox, I. J. (2021). Tracking covid-19 using online search. NPJ digital medicine, 4(1), 17.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of google flu: traps in big data analysis. Science, 343(6176), 1203–1205.

Liu, Z., Jiang, Z., Kip, G., Snigdha, K., Xu, J., Wu, X., . . . Schultz, T. (2022). An infodemiological framework for tracking the spread of sars-cov-2 using integrated public data. Pattern Recognition Letters, 158, 133–140.

Nagao, S., Takeda, F., & Tanaka, R. (2019). Nowcasting of the us unemployment rate using google trends. Finance Research Letters, 30, 103–109.

Petropoulos, A., Siakoulis, V., Stavroulakis, E., Lazaris, P., & Vlachogiannakis, N. (2022). Employing google trends and deep learning in forecasting financial market turbulence. Journal of Behavioral Finance, 23(3), 353–365.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using google trends. Scientific reports, 3(1), 1684.

Suhoy, T. (2009). Query indices and a 2008 downturn: Israeli data (Tech. Rep.). Bank of Israel.

WB. (2024). World Development Indicators. Washington, D.C., US. (World Bank. Retrieved from: https://databank.worldbank.org/source/world-development-indicators [Accessed February 1, 2024])

Zola, P., Ragno, C., & Cortez, P. (2020). A google trends spatial clustering approach for a worldwide twitter user geolocation. Information Processing & Management, 57(6), 102312.