CARMA 2024

6th Int. Conf. on Advanced Research Methods and Analytics
Universitat Politècnica de València, València, 2024
DOI: https://doi.org/10.4995/CARMA2024.2024.17158

Google Trends Forecasting of Youth Unemployment

Nathan de Bruijn, Fons Wijnhoven , Robin Effing
Behavioural, Management and Social Sciences, University of Twente, The Netherlands.

How to cite: De Bruijn N.; Wijnhoven F.; Effing R. 2024. Google Trends Forecasting of Youth Employment. In: 6th International Conference on Advanced Research Methods and Analytics (CARMA 2024). Valencia, 26-28 June 2024. https://doi.org/10.4995/CARMA2024.2024.17158

Abstract
The forecasting field has been using the surge in big data and advanced computational capabilities. This article discusses the methodological issues of Google Trends (GT) data reliability and forecasting validity for youth unemployment forecasts. We demonstrate the problems with static GT forecasting procedures and show a 44% increase in forecasting accuracy by applying time-varying model respecification forecasting.

Keywords: *Forecasting; time series; rolling window; expanding window; unemployment; google trends*

1. Introduction

The complexity of using GT data in forecasting is reflected in the challenge of selecting suitable keywords from an array of millions of potential keywords (Varian, 2014). This complexity is amplified since GT data structurally changes over time (Behnen et al., 2020). This makes predicting with GT from both data reliability and prediction validity perspectives challenging. Therefore, our research question is "How can GT data reliably be used for generating valid unemployment predictions?" While answering this question, this study aims at contributing to insights on managing GT data reliability and forecasting in time-variant contexts.

In this article we first discuss literature related to benefits and drawbacks of using GT data. Subsequently, the methodology section outlines the research design. The results section presents our forecasts. Finally, we present our conclusions and discuss their implications.

2. Literature review

Choi & Varian (2012) demonstrated that the popularity of Google searches like "apply for unemployment" are useful in forecasting future unemployment. Similarly, Ginsberg et al. (2009) launched a tool called Google Flu Trends to forecast flue occurrences, but Google Flu Trends faced notable criticism when it overestimated doctor visits by a factor of more than two (Lazer et al., 2014). One of the causes was Google Flu Trends' continuous search for the most correlated keywords, without theorizing ex-ante which keywords (i.e., predictor variables) are appropriate (Lazer et al., 2014). Moreover, algorithmic changes caused Google Flu Trends to become less accurate over time and GT was phased out by Google (Lazer et al., 2014).

The use of GT data does come with several advantages however. GT data is fully anonymized and data collection occurs without any effort from the user. Users may not even be aware of data collection, ensuring that the recorded data is unobtrusive and reflects natural behaviour (McLaren & Shanbhogue, 2011). Other advantages include the real-time availability of pre-processed data at no cost (Zhu et al., 2012). Nevertheless, pre-processing by Google does come with sampling error (Cebrián & Domenech, 2022). GT data is also criticized to be unreliable, due to the homogeneity of internet users having an influence on keyword popularity. Also, individuals may lack internet access or use alternative sources of search, resulting in GT coverage bias (Cebrián & Domenech, 2022).

Recent literature also discussed the inconsistency of GT data. For example, GT data for the same keyword, during the same period, may be different when again collected tomorrow (Cebrián & Domenech, 2022). Eichenauer et al. (2021) attribute this inconsistency to sampling variation, which is more noticeable for less popular keywords and smaller regions due to smaller samples. Furthermore, GT data has the tendency to structurally change over time (Behnen et al., 2020) giving rise to the issue of parameter instability in forecasting. Furthermore, spurious correlations between GT data and phenomena that need to be forecasted are found. For example, GT data for a popular drink was highly correlated to housing sales in the USA (Tran et al., 2017).

Even with these limitations, GT provides researchers with a large dataset of user behavior related to real world developments, useful if the data reliability and forecasting validity issues are handled well.

3. Methodology

3.1. Unemployment forecasting with GT

There are many studies that leveraged GT data to forecast unemployment focused on a single keyword. Examples being: D'Amuri & Marucci (2017) who used the keyword "jobs", Mclaren & Shanbogue (2011) who used "jobseeker's allowance", Simionescu & Cifuentes-Faura (2022) who used "unemployment", Naccarato et al. (2018) who used "job offers", Fondeur & Karamé (2013) who used "employment", and Vicente et al. (2015) who used "job offer". Other studies leveraged multiple keywords. For example, Tuhkuri (2016) created an index by averaging over thirteen keywords with weights based on a Google search volumes. However, the aforementioned studies relied on intuition, and no formal keyword selection techniques were used. Other studies adopted formal techniques for keyword selection, like Borup & Schutte's (2020) regularization approach and Singhania & Kundu (2021) who inputted over 500 potential keywords to a neural network. Also, Li et al. (2015) applied dimension reduction techniques to select keywords.

3.2. Data collection

The GT dataset we use for this study is the GT search volume index (SVI). A low SVI indicates a low search volume for the search keyword. Either weekly or monthly data can be obtained from GT. GT data spanning from April 2008 till December 2022 is collected for this study, which is in line with the time length of the unemployment dataset that we gained from the Dutch Census Office CBS. To deal with data invalidity, this study started from a domain ontology (Guizzardi et al., 2022), to then only select keywords based on literature and economic reasoning. This was carried out by the following steps.

1. To ensure that keywords are representative of the total number of unemployment related searches, we estimated the monthly search volume with the Google Ads Keyword Planner. Keywords with a low monthly search volume were considered not representative.
2. 26 studies that forecasted unemployment with GT data were studied to find the keywords they used. 329 keywords were obtained and checked for fitting into the domain ontology of unemployment. This led to the identification of 20 main themes, six of the prominently used are: (1) job search, (2) unemployment interest, (3) employment agency, (4) job platform, (5) unemployment benefits, and (6) unemployment claims. These six themes were used to find their Dutch equivalents. Additionally, Dutch keywords were taken from (Te Brake, 2017). Keywords with less than 1,000 monthly searches were removed. This resulted in 58 remaining keywords.
3. These 58 keywords were used to prompt Google's algorithm to return closely related keywords. The search volumes of the closely related keywords were also checked and after doing so another 21 keywords were added, resulting in a total of 79 keywords that accounted for 2,799,400 monthly Google searches.
4. The GT data for each keyword was obtained at 12 different moments across 9 days to reduce sampling variation. This resulted in 12 datasets for each keyword, summing to a total of 948 datasets.
5. The mean correlation between the 12 GT data samples for the same keyword was checked. All keywords with a mean correlation lower than .90 were dropped. Consequently, 63 keywords remained fit for analysis. For each of the 63 keywords that remained, the 12 datasets obtained for these keywords were averaged, as suggested by Eichenauer et al. (2021).

3.3. Analysis

Two procedures exist for out-of-sample forecasting; the fixed-origin and the rolling-origin (Hewamalage et al., 2023). While the fixed-origin procedure for out-of-sample forecasting has long been used, the rolling-origin procedure for out-of-sample forecasting has become the favorite choice. With the fixed-origin procedure, the complete dataset is divided into a training set and test set, and this division remains constant, i.e., all following predictions being based on the same training window. For the rolling window approach, the window size may be extended (expanding window) or the window size may remain the same but each prediction of a following period may be on a later start of the training window (rolling window). For the actual predictions two approaches are commonly used (Hewamalage et al., 2023): 1) updating the forecasting model by feeding it with new data and thus without any change of the prediction model, 2) using new data to recalibrate the forecasting model, i.e., for example improving parameter values. This study introduces a third approach, the inclusion of other variables and relations in the prediction model, so-called re-specification, to find a prediction model that may predict better than the previous one. Disadvantageous to model recalibration and respecification is the computational power that is needed. However, thanks to increased computational power, model recalibration has already become a common practice in forecasting (Hewamalage et al., 2023). Re-specification now becomes a more relevant option to prediction modeling because of the wide availability of millions of potential social media and or GT predictor variables that have the tendency to structurally change over time. Consequently, the selection of suitable variables emerges as a more pressing concern than further complicating forecasting models..

In our study, out-of-sample forecasts for the period October 2016 till December 2022 are produced 49 times in this study with each of the out-of-sample forecasting procedures (i.e., rolling window with model recalibration (RC), expanding window with model recalibration (EC), rolling window with model respecification (RS), and expanding window with model respecification (ES)), each time using a different window size ranging from 48 to 96. After averaging over the 49 forecasts for the out-of-sample October 2016 till December 2022, a single robust forecastis obtained for each out-of-sample forecasting procedure. For the best performing forecasting procedure, the multiple linear regression model that is used is extended with an autoregressive (AR) component to add a lagged version of the dependent variable (Hyndman & Athanasopoulos, 2018).

4. Results

The first part of this section presents the senstivity of forecasting accuracy to the window size. Subsequently, the forecasts produced forecast are evaluated . Finally, the AR(1) is added.
The out-of-sample period used for each window size is October 2016 till December 2022, while the in-sample periods vary between 1-10-2008 and 1-12-2012. Figure 1, 2, 3, and 4 give the results and indicate that the forecast error, and thus accuracy, is sensitive to the window size. However, the relationship between window size and forecast error is complex and can hardly be captured with an equation. Moreover, for each out-of-sample forecasting procedure, the relationship between window size and forecast error is different. Not only does the pattern differ, but the magnitude of the effect is also different. This has led to the decision to average over the forecasts obtained from the 49 different windows that were used in training the model.

Figure 1-4 summarize that the out-of-sample forecasting procedures relying on model re-specification result in more accurate forecasts than those relying on model recalibration both for the rolling window (Figure 1 vs 3) and the expanding window (Figure 2 vs 4).

Table 1 further explains only gives weak evidence that when model re-specification is used, the rolling window will produce more accurate results than the expanding window. Given that the error diagnostics of the best performing forecasting procedure (RS with window size of 48) indicated that some information is missing, an autoregressive component with a lag order of 1 is added, referred to as AR(1). Using an AR(1) component means that last month's youth unemployment rate is used to predict this month's youth unemployment rate. The forecast is now produced for a larger out-of-sample than previously used, ranging from October 2012 till December 2022.

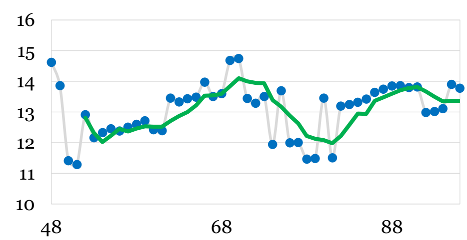

Figure 1. Forecast error rolling re-calibration to window size

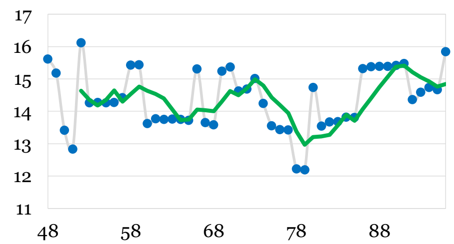

Figure 2. Expanding calibration forecast error

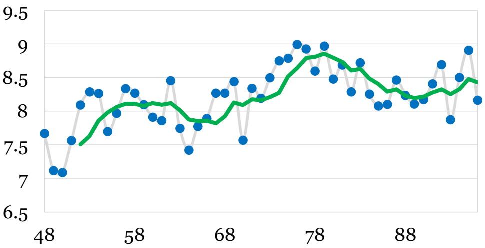

Figure 3. Forecast error rolling window re-specification

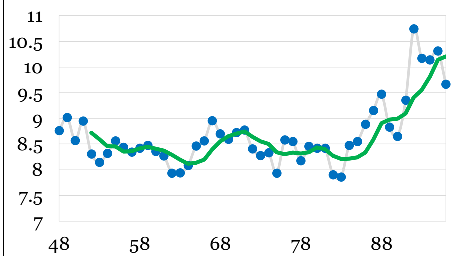

Figure 4. Forecasting error expanding window re-specification

Note: MAPE is presented on the y-axis. The MAPE for a certain window size is calculated from a sample of 75 forecasts and their corresponding errors. Window size is presented on the x-axis. The green line in the graphs represents the moving average of the MAPE. thick line concerns moving average.

The autocorrelation of the forecast errors is reduced when the AR(1) component is added to the multiple linear regression model, resulting in no significant autocorrelation at any lag. When the AR(1) component is added, the forecasting errors are also less dispersed and lower than when the AR(1) component is not added. The assumptions of multiple linear regression are fulfilled to a greater extent when an AR(1) component is used in addition to the GT variables. The one month ahead forecasts of the youth unemployment rate are more accurate when supplementing the multiple linear regression model with an AR(1) component, with MSE being .4375 as opposed to .8079. Similarly, RMSE is substantially down from .8988 to just .6615. Finally, MAPE is also lower when the AR(1) component is used in addition to the GT variables, decreasing from 6.75% to just 4.73%.

Figure 5 shows that the forecasts align well with the actual youth unemployment rate, even for the period characterized by COVID-19. The overall fit of the forecasting model with reality has an Adjusted R-Squared of 91.33%. Although the multiple regression model, relying on the RS procedure for out-of-sample forecasting, is relatively unbiased and accurate this section revealed that it is important to supplement GT data with additional data, like the AR(1) component.

5. Conclusions and Discussion

This study reminds forecasting literature of the limitations inherent to GT data. Moreover, this study found that forecasts relying on solely GT data are substantially improved when additional information, like an autoregressive component, is added. The findings also contribute to literature by: (1) Raising awareness on the importance of picking the correct out-of-sample forecasting procedure, and (2) demonstrating that forecasts can be improved by using a different out-of-sample forecasting procedure. The results of our study also aligns with Shen et al. (2020), revealing that a larger window size may lead to lower forecasting accuracy, both for the rolling window and the expanding window.

Table 1. Error metrics

	MSE (n=75)	RMSE (n=75)	MAPE (n=75)
ES	.7302	.8545	7.807%
EC	1.936	1.391	13.34%
RS	.7816	.8841	7.213%
RC	1.589	1.260	11.94%

Note: One month ahead forecasts averaged over windows from 48 till 96 months.

We find that model re-specification substantially improves the accuracy when out-of-sample forecasting the youth unemployment rate. Compared to model recalibration, model re-specification yields 44% more accurate forecasts of the youth unemployment rate. This finding is supported with 99% of confidence. Consequently, the dominance of model re-specification, as opposed to model recalibration may be generalizable for both the rolling window and the expanding window procedure for out-of-sample forecasting.

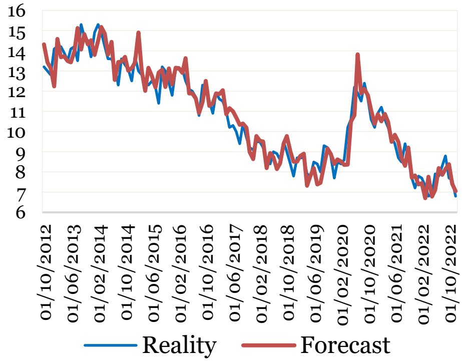

Figure 5. Best youth unemployment GT forecasts aligned with reality & AR(1) added. Note: Rolling window size of 48.

We acknowledge some limitations in our work. First, model re-specification requires large computational power, especially when more complex modelling techniques are used. For example, an autoregressive integrated moving average model takes around 60 times longer to process out-of-sample forecasts than a multiple linear regression model when using model respecification. Second, this study merely established correlation, and not causation, between GT data and the youth unemployment rate. It could be the case that the youth unemployment rate is explaining GT data more strongly than the other way around. Third, the keywords that are used could be subject to noise. For example, keywords like “werkloosheid” (unemployment) do not solely reflect searches done by the unemployed. Rather, searches for this keyword could simply reflect an interest in the current state of the economy. Fourth, this study introduced coverage bias by using merely Dutch keywords, therefore excluding individuals that don’t speak Dutch. Fifth, this study only forecasted one month ahead, limiting the practical value.

Inspired by the limitations, there are various future research suggestions. Specifically we suggest that the selection of keywords/variables could be fully automated. For example, ChatGPT could be prompted to return keywords related to a certain domain ontology. Subsequently, Google Tends data for these keywords could be obtained automatically.

References

Behnen, P., Kessler, R., Kruse, F., Gómez, J. M., Schoenmakers, J., & Zerr, S. (2020). Experimental Evaluation of Scale, and Patterns of Systematic Inconsistencies in Google Trends Data. *Communications in Computer and Information Science*, *1323*, 374–384. https://doi.org/10.1007/978-3-030-65965-3_25/TABLES/4

Borup, D., & Schütte, E. C. M. (2020). In Search of a Job: Forecasting Employment Growth Using Google Trends. *Journal of Business & Economic Statistics*, *40*(1), 186–200. https://doi.org/10.1080/07350015.2020.1791133

CBS. (2023). *Arbeidsdeelname en werkloosheid per maand*. https://opendata.cbs.nl/#/CBS/nl/dataset/80590ned/table?dl=770B2

Cebrián, E., & Domenech, J. (2022). Is Google Trends a quality data source? *Applied Economics Letters*, *30*(6), 811–815. https://doi.org/10.1080/13504851.2021.2023088

Chatfield, C., & Xing, H. (2019). *The analysis of time series: an introduction with R* (7th ed.). CRC Press.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, *88*(special issue SI), 2–9.

D'Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, *33*(4), 801–816. https://doi.org/10.1016/j.ijforecast.2017.03.004

Eichenauer, V. Z., Indergand, R., Martínez, I. Z., & Sax, C. (2021). Obtaining consistent time series from Google Trends. *Economic Inquiry*, *60*(2), 694–705. https://doi.org/10.1111/ecin.13049

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. https://doi.org/10.1038/nature07634

Hewamalage, H., Ackermann, K., & Bergmeir, C. (2023). Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*, *37*(2), 788–832. https://doi.org/10.1007/s10618-022-00894-5

Hyndman, R., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.com.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506

Li, X., Shang, W., Wang, S., & Ma, J. (2015). A MIDAS modelling framework for Chinese inflation index forecast incorporating Google search data. *Electronic Commerce Research and Applications*, *14*(2), 112–125. https://doi.org/10.1016/j.elerap.2015.01.001

McLaren, N., & Shanbhogue, R. (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin*, *51*(2), 134–140. http://econpapers.repec.org/RePEc:boe:qbullt:0052

Mulero, R., & García-Hiernaux, A. (2021). Forecasting Spanish unemployment with Google Trends and dimension reduction techniques. *SERIEs*, *12*(3), 329–349. https://doi.org/10.1007/s13209-021-00231-x

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

Shen, Z., Zhang, Y., Lu, J., Xu, J., & Xiao, G. (2020). A novel time series forecasting model with deep learning. *Neurocomputing*, *396*, 302–313. https://doi.org/10.1016/j.neucom.2018.12.084

Simionescu, M., & Cifuentes-Faura, J. (2022). 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

Singhania, R., & Kundu, S. (2021). Forecasting the United States Unemployment Rate by Using Recurrent Neural Networks with Google Trends Data. *International Journal of Trade, Economics and Finance*, *11*(6). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3801209

Te Brake, G. (2017). *Unemployment? Google it! Analyzing the usability of Google queries in order to predict unemployment*. Universitat de Barcelona.

Tran, U. S., Andel, R., Niederkrotenthaler, T., Till, B., Ajdacic-Gross, V., & Voracek, M. (2017). Low validity of Google Trends for behavioral forecasting of national suicide rates. *PloS One*, *12*(8), e0183149–e0183149. https://doi.org/10.1371/journal.pone.0183149

Tuhkuri, J. (2016). *Forecasting unemployment with google searches* (35). https://www.econstor.eu/handle/10419/201250

Varian, H. R. (2014). Big data: New tricks for econometrics. *The Journal of Economic Perspectives*, *28*(2), 3–27.

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