CARMA 2024

6th Int. Conf. on Advanced Research Methods and Analytics
Universitat Politècnica de València, València, 2024
DOI: https://doi.org/10.4995/CARMA2024.2024.17783

Economic forecasting with non-specific Google Trends sentiments: Insights from US Data

Sami Diaf[1], Florian Schütze[2]

[1]Department of Socioeconomics, University of Hamburg, Germany. [2]Faculty of Economics and Social Sciences, Helmut Schmidt University, Germany.

How to cite: Diaf, S.; Schütze, F. 2024. Economic forecasting with non-specific Google Trends sentiments: Insights from US Data. In: 6th International Conference on Advanced Research Methods and Analytics (CARMA 2024). Valencia, 26-28 June 2024. https://doi.org/10.4995/CARMA2024.2024.17783

Abstract

The influence of specific Google Trends search queries measuring various sentiments on economic performance and stock markets has been extensively documented and used for many purposes. This paper examines the predictive power of queries measuring non-specific sentiment on key macroeconomic variables when linked to a comprehensive sentiment dictionary. The analysis shows that non-specific sentiments do not improve the forecasting quality of the US economy as a whole, except for unemployment, which was found to be predictable for all sentiments. Consequently, the authors suggest that economic-related sentiments with carefully selected words should be used in Google Trends search queries to improve predictive performance. However, if a socio-cultural analysis is to be performed, non-specific sentiments would be suggested, as they can be predicted by the real economic time series of unemployment.

Keywords: *Sentiment analysis; Google Trends and Search Engine data; Web scraping; Internet econometrics; Forecasting and nowcasting.*

1. Introduction

Social listening is a regular source of data used by economists to test further hypotheses beyond the available economic aggregates and other indices. The sentiment of economic agents could be used as a proxy to predict the evolution of economic time series.

Affuso & Lahtinen (2019) showed that negative sentiment among Twitter users has a greater impact on stock returns than positive sentiment. Eugster & Uhl (2024) established an improvement in the accuracy of an inflation forecast using a self-generated sentiment index based on newspaper articles. Rambaccussing & Kwiatkowski (2020) also used sentiment analysis of newspaper articles and found it useful for forecasting unemployment and output. Sharpe et al. (2017) analyzed optimistic and pessimistic sentiment in Federal Reserve Board forecasts and discovered that these sentiments can predict both GDP growth and unemployment.

Since the above literature has shown that sentiment measured in various media has an impact on the real economy, this paper will investigate whether this can be applied to Google Trends and non-specific sentiment. There are examples in the literature that this works with specific sentiment words.

Broachado (2020) created a Google Sentiment Index that measures the overall polarity about the economy and shows short-term predictive ability regarding the stock market. Borup & Schütte (2022) show that labor market forecasts can be improved by using specific labor market searches on Google. While these papers use Google Trends with keywords, it has also been shown that economic uncertainty sentiment, measured with economic topics instead of keywords, has an impact on the economy (Schütze, 2020; Schütze, 2022). Donadelli and Gerotto (2019) showed that an increase in search queries for non-macro-based topics had a negative impact on economic time series. However, these topics were related to health, environment, security, and politics, which means that they are also specific in some sense.

Therefore, this paper explores the possibility of improving the forecasting quality of economic time series by analyzing non-specific sentiments in Google Trends, with the goal of improving the forecasting quality of macroeconomic time series. For this purpose, the sentiment dictionary of Loughran & McDonald (2011), which contains 8 different sentiment categories, is used, although its scope is not limited to the economic context, but the words measure the basic sentiment in a publication or in Google Trends. Thus, this application investigates the duality of specific/non-specific sentiments and their predictive abilities in the economic domain.

The results show that non-specific sentiments do not increase the overall explanatory power of macroeconomic variables, but only two out of eight sentiments have an impact on some of the economic time series. This shows that the approach with specific sentiment words is justified, as Google Trends time series with non-specific words do not have the same predictive quality as specific words. Moreover, unemployment is the only aggregate that has a predictive quality for all 8 sentiments, meaning that it captures the interest of most economic agents as translated by their search queries.

2. Methodology

Loughran & McDonald's (2011) sentiment word list has been used for many sentiment analysis exercises in the social sciences and, more recently, on social listening sources (Google Trends). It contains eight (8) sentiment categories: Negative, Positive, Uncertainty, Litigious, Strong_Modal, Weak_Modal, Constraining, and Complexity. Table 1 shows the different sentiment categories and the number of words within each category. It also shows the number of words that resulted in a complete time series of Google Trends search queries. Approximately 10% of the words did not result in a complete Google Trends query.

For each word in the sentiment word list, a Google Trends query was run using the R package "gtrendsR" (Massicotte et al., 2016). A total of 4,194 queries were run, of which 3,740 resulted in a Google Trends time series. This means that for each of the 3740 different queries, a Google Trends time series for the US was downloaded for the period 01/2004 to 12/2023.

A total of eight Google Trends sentiment indices were created from the words in each sentiment category. Four different "weights" were applied: 1. The average of all Google Trends time series within a category; 2. The weighted average of all Google Trends time series. The weight is determined by the relative frequency with which each word occurs within a category (Loughran & McDonald, 2011); 3. A principal component analysis was performed on all words in each category. The third weight is the first principal component of the analysis, and the fourth weight is the second principal component. The first explains approximately 40% of the data, the second approximately 20% in each sentiment category.

Table 1. The number of words in the sentiment dictionary from Loughran & McDonald (2011) and the number of Google queries available to match them. Source: Own calculation.

	Negative	**Positive**	**Uncertainty**	**Litigious**	**Strong_Modal**	**Weak_Modal**	**Constraining**	**Complexity**	**Sum**
Words	2355	354	297	905	19	27	184	53	4194
Queries	2195	289	294	727	17	27	149	42	3740

Figure 1 shows the eight different sentiment categories and the four different weightings. Note that positive and negative sentiment show a similar trend in the two different weightings. Only the principal component analysis shows an opposite trend for negative and positive sentiment. The same is true for the distinction between Weak_Modal and Strong_Modal sentiment. The chart shows that the approach of testing different weightings can be useful, as there are significant differences.

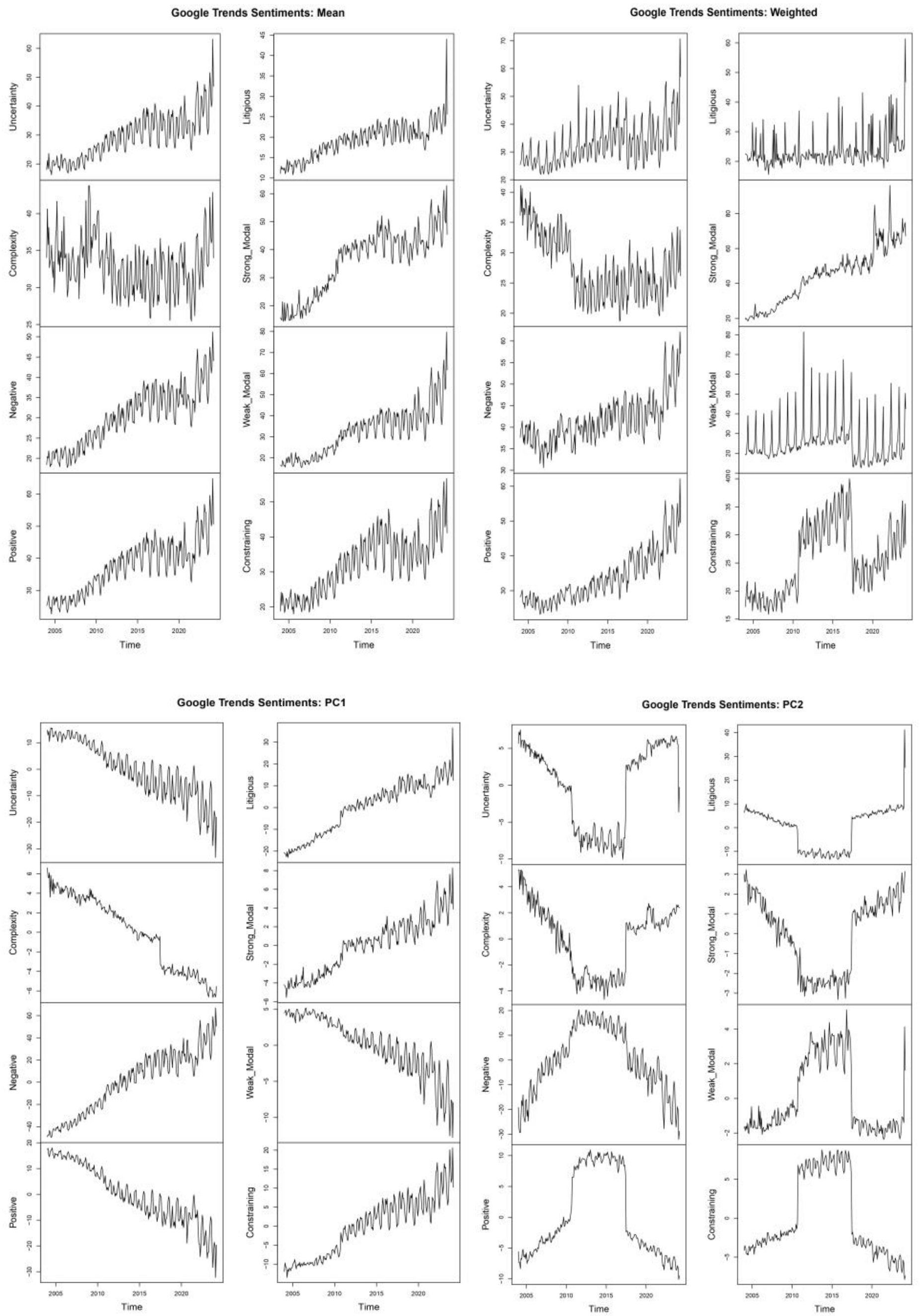

Figure 1. Google Trends Sentiments: different weightings. Source: Own calculation.

The Granger causality analysis used all 8 different sentiments with the 4 different weights, resulting in 32 time series, as well as 4 economic monthly time series: personal consumption expenditures (U.S. Bureau of Economic Analysis, 2024), consumer prices (OECD, 2024b), monthly unemployment rates (OECD, 2024c), and industrial production (OECD, 2024a). The time period is from 01/2004 to 12/2023. The research question to be tested here is whether the underlying sentiment, as measured by Google Trends, has predictive power for real economic time series. At the same time, it is also tested whether the real economic time series can predict the sentiment. It is important to note that Granger causality is not causality in the classical sense, but only shows that one time series has predictive power with respect to another time series.

All series are seasonally adjusted by creating a dummy variable for each month. The seasonally adjusted series is the residual series from a regression with the original series as the endogenous variable and the months without constants as the exogenous variable. The Augmented Dickey-Fuller test was then used to determine whether the time series were I(0) or I(1). The method of Toda & Yamamoto (1995) was used for the Granger causality analysis because it is robust to different stationary orders of the time series. Since different time series are analyzed, the following combination could occur: Sentiment time series = I(0) and industrial production= I(1).

The lag selection for the VAR model was done using Akaike's information criterion, with a maximum lag length of 12. A time series is assumed to be Granger causal for another time series if the null hypothesis that there is no Granger causality can be rejected at the 5% significance level. The Wald test used in Granger causality analysis is based on the rather simple premise of comparing the performance of a restricted model Y, which excludes X, with an unrestricted model for Y, which includes X.

3. Results

The analysis of Granger causality in the direction of sentiment towards the economic time series in Table 2 shows that sentiment does not have much predictive power. The null hypothesis that sentiment Granger-causes consumer prices cannot be rejected with $\alpha<5\%$ for a single sentiment time series. Industrial production can only be Granger-caused by two sentiment time series, the first principal component of the positive sentiment category and the weighted Strong_Modal sentiment. Personal consumption expenditures can be Granger-caused by 3 sentiment time series: The average of the Litigious sentiment, the first principal component of the Litigious sentiment, and the average of the Strong_Modal sentiment. Unemployment can also be Granger causally explained by three sentiment time series. In this case, it is the first principal component of the Complexity Sentiment. In addition, the average of the Litigious Sentiment and the average of the Strong_Modal Sentiment can Granger causally explain unemployment.

Therefore, there are sentiments that in some cases have increased predictive performance. These include the Strong_Modal and Litigious categories. Given that 32 different sentiment time series were used, with different weights for each economic variable, it is expected that 5% of the 32 models will reject H0 of the Granger test by chance. On average, this would be 1.6 models. On average, 2 models per economic variable lead to a rejection of H0. Given this, the authors conservatively assume that Google Trends sentiment has no fundamental Granger causality on economic time series.

Table 2. p-values of the Granger causality analysis, both directions. In red and bold: α<5% when H0 is rejected. Source: Own calculation.

	H0: Sentiment does not Granger cause one of the economic time series				H0: An economic time serie does not Granger cause one of the sentiments			
	CPI	Ind. Prod.	PCE	Unemp.	CPI	Ind. Prod.	PCE	Unemp.
UncertaintyMean	0,890	0,455	0,849	0,562	0,547	0,762	0,096	**0,006**
UncertaintyWeight	1,000	0,672	0,665	0,476	0,852	0,328	0,129	0,107
UncertaintyPC1	0,913	0,239	0,574	0,309	0,688	0,269	**0,041**	**0,000**
UncertaintyPC2	0,957	0,132	0,938	0,112	0,929	0,814	0,999	0,088
ComplexityMean	0,519	0,737	0,816	0,535	0,099	0,059	0,464	**0,003**
ComplexityWeight	0,986	0,436	0,831	0,485	0,162	0,242	0,103	**0,000**
ComplexityPC1	0,688	0,709	0,876	**0,030**	0,214	0,932	0,805	0,136
ComplexityPC2	0,108	0,614	0,968	0,275	0,389	0,802	0,730	0,154
NegativeMean	0,206	0,470	0,067	0,160	0,535	0,431	0,358	0,117
NegativeWeight	0,753	0,849	0,341	0,500	0,103	0,803	0,888	0,632
NegativePC1	0,103	0,484	0,052	0,126	0,865	0,231	0,132	**0,022**
NegativePC2	0,963	0,371	0,237	0,148	0,788	0,552	0,233	0,051
PositiveMean	0,421	0,081	0,155	0,178	0,136	0,198	0,174	**0,027**
PositiveWeight	0,913	0,102	0,192	0,091	0,456	0,254	0,536	0,056
PositivePC1	0,489	**0,040**	0,232	0,163	0,285	0,130	0,304	**0,006**
PositivePC2	0,551	0,739	0,872	0,299	0,836	0,501	0,866	**0,027**
LitigiousMean	0,101	0,152	**0,017**	**0,033**	0,362	0,516	0,388	0,473
LitigiousWeight	0,298	0,366	0,136	0,367	**0,023**	0,563	0,661	**0,004**
LitigiousPC1	0,207	0,186	**0,028**	0,106	0,843	0,724	0,449	0,214
LitigiousPC2	0,466	0,564	0,712	0,201	0,783	0,849	0,472	0,120
Strong_ModalMean	0,328	0,080	0,846	0,608	0,652	0,760	0,748	**0,006**
Strong_ModalWeight	0,172	**0,000**	**0,023**	**0,000**	0,883	0,340	**0,033**	0,383
Strong_ModalPC1	0,693	0,187	0,674	0,418	0,858	0,923	0,778	**0,030**
Strong_ModalPC2	0,556	0,885	0,943	0,488	0,732	0,236	0,472	0,244
Weak_ModalMean	0,941	0,572	0,403	0,424	0,601	0,701	0,296	**0,014**
Weak_ModalWeight	0,961	0,922	0,928	0,844	0,777	0,964	0,997	0,772
Weak_ModalPC1	0,970	0,510	0,330	0,368	0,127	0,264	0,166	**0,001**
Weak_ModalPC2	0,628	0,995	1,000	0,722	0,769	0,908	0,994	0,303
ConstrainingMean	0,745	0,393	0,620	0,535	0,408	0,662	0,116	**0,007**
ConstrainingWeight	0,996	0,493	0,964	0,657	0,454	0,722	0,561	0,251
ConstrainingPC1	0,900	0,373	0,442	0,288	0,624	0,826	0,143	**0,008**
ConstrainingPC2	0,922	0,388	0,990	0,302	0,977	0,999	0,961	0,077

When the Granger causality sequence is reversed, it turns out that consumer prices can only Granger causally explain the average of the litigious. Industrial production cannot Granger causally explain a single sentiment. Consumer spending has predictive power for the first principal component of uncertainty. In addition, consumer spending can explain the average of Strong_Modal. Unemployment can Granger causally explain a total of 15 out of 32 sentiments. In each sentiment category, there is at least one weighting scheme that is Granger causally influenced by unemployment.

Contrary to the previous statement that Google Trends sentiment has no predictive quality for economic variables, it is now apparent that US unemployment has predictive quality for Google Trends sentiment. With this discovery, the focus of this paper changes from a (macro)economic analysis to a sociological analysis. It turns out that general, non-economic sentiments have no influence on future economic development. On the other hand, a change in unemployment affects all categories of sentiment and these can be better predicted. It seems that the socio-cultural influence of unemployment on the mood of the population is very strong and pronounced.

4. Conclusion

The initial working hypothesis that unspecific sentiment, as measured by Google Trends, can be used to improve the forecasting performance of monthly economic time series is likely to be rejected. It turns out that only two sentiments show improved forecasting performance, while the remaining do not. In contrast, a reverse Granger causality analysis shows that unemployment has an impact on each sentiment category. This is confirmed by different weighting schemes. If this predictive quality is equated with an influence from one time series to another, it becomes clear that unemployment has a strong influence on the polarization of society. Even without this equation, it is obvious that there must be a correlation, since lack of work can be an existential experience that also strongly polarizes individuals.

Furthermore, this paper shows that the results of previous applications of sentiment measurement cannot be generalized. For example, non-specific sentiment uncertainty does not affect economic variables, but specific economic sentiment "uncertainty" does.

The finding that unemployment has a Granger causal effect on all sentiments shows that the Google Trends approach is promising. Theoretically, a change in unemployment should capture the interest of society, which in turn should lead to more Google searches. However, this finding also reinforces the statement that only (economic) specific sentiments should be used if the predictive power of these sentiments is to be increased. This can be done by choosing an appropriate topic or by concatenating words. Another area of research would be to investigate the specific influence of unemployment on sentiments, and whether this is the case across regions. For example, urban regions with an affinity for the Internet may be more affected than suburban or rural regions.

References

Affuso, E., Lahtinen, K.D. (2019). Social media sentiment and market behavior. Empirical Economics 57, 105–127.

Borup, D., & Schütte, E. C. M. (2022). In search of a job: Forecasting employment growth using Google Trends. Journal of Business & Economic Statistics, 40(1), 186-200.

Brochado, A. (2020). Google search based sentiment indexes. IIMB Management Review, 32(3), 325-335.

Donadelli, M., and Gerotto, L. (2019). Non-macro-based Google searches, uncertainty, and real economic activity. Research in International Business and Finance, 48, 111-142.

Eugster, P., and Uhl, M. W. (2024). Forecasting inflation using sentiment. Economics Letters Economics Letters, Volume 236, 111575.

Massicotte, P., Eddelbuettel, D. and Massicotte, M. P. (2016). Package 'gtrendsR', R package.

Loughran, T. and McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. The Journal of Finance, 66 (1), 35-65.

OECD (2024a). Industrial production (indicator), doi: 10.1787/39121c55-en (Accessed on 14 February 2024).

OECD (2024b). Inflation (CPI) (indicator), doi: 10.1787/eee82e6e-en (Accessed on 14 February 2024)

OECD (2024c), Unemployment rate (indicator), doi: 10.1787/52570002-en (Accessed on 14 February 2024)

Rambaccussing, D. and Kwiatkowski, A. (2020). Forecasting with news sentiment: Evidence with UK newspapers. International Journal of Forecasting, 36 (4), 1501-1516.

Schütze, F. (2020), Google Trends Topic-Based Uncertainty: A Multi-National Approach, CARMA 2020 - 3rd International Conference on Advanced Research Methods and Analytics, Valencia, Spain, 8 – 9 July 2020, pp. 191-199.

Schütze, F. (2022), The demand side of information provision: Using multivariate time series clustering to construct multinational uncertainty proxies, CARMA 2022 - 4rd International Conference on Advanced Research Methods and Analytics, Valencia, Spain, 29 June – 1 July 2022, pp. 155–163.

Sharpe, S A., Sinha, N. R., and Hollrah, C. A. (2017). What's the Story? A New Perspective on the Value of Economic Forecasts. Finance and Economics Discussion Series 2017-107, Board of Governors of the Federal Reserve System (U.S.).

Toda, H. Y. and Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. Journal of Econometrics, 66 (1-2), 225–250.

U.S. Bureau of Economic Analysis (2024). Personal Consumption Expenditures [PCE], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/PCE (Accessed on 14 February 2024).