

Karaman Örsal, Deniz Dilan; Sturm, Silke

Working Paper

Computing Consumer Sentiment in Germany via Social Media Data

Hamburg Discussion Papers in International Economics, No. 7

Provided in Cooperation with:

University of Hamburg, Department of Economics, Senior Lectureship in International Economics

Suggested Citation: Karaman Örsal, Deniz Dilan; Sturm, Silke (2021) : Computing Consumer Sentiment in Germany via Social Media Data, Hamburg Discussion Papers in International Economics, No. 7, University of Hamburg, Chair of International Economics, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/229183>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Computing Consumer Sentiment in Germany via Social Media Data

Deniz Dilan Karaman Örsal

Silke Sturm

Hamburg Discussion Papers in International Economics [No. 7]

University of Hamburg

Chair of International Economics

University of Hamburg

Chair of International Economics

The Chair of International Economics is located at the University of Hamburg. The research focuses on the areas of foreign trade, European economic integration and migration.

The Discussion Paper Series is designed to enable internal and external researchers as well as excellent students to make their research at the chair publicly available. The aim is to contribute to the current debate of policy and research relevant topics. The views expressed in the paper represent the opinion of the authors only. For any questions or comments, please directly contact the author.

Editor:

Universität Hamburg

Chair of International Economics

Prof. Dr. Thomas Straubhaar

Von-Melle-Park 5

20146 Hamburg

<https://www.wiso.uni-hamburg.de/en/fachbereich-vwl/professuren/straubhaar.html>

Deniz Dilan Karaman Örsal

Universität Hamburg

Von-Melle-Park 5, 20146 Hamburg

+49-40-42838-8151

deniz.oersal@uni-hamburg.de

ISSN 2625 - 7513 (online)

Computing Consumer Sentiment in Germany via Social Media Data

Deniz D. Karaman Örsal[†], Silke Sturm[‡]

Universität Hamburg

January 29, 2021

Preliminary draft

Please do not quote without author's permission.

Abstract

Survey-based consumer confidence indicators are mostly reported with a delay and are a result of time consuming and expensive consumer surveys. In this study, to measure the current consumer confidence in Germany, we develop an approach, in which we compute the consumer sentiment using public *Tweets* from Germany. To achieve this goal we develop a new sentiment score. To measure the consumer sentiment, we use text-mining tools and public *Tweets* from May 2019 to August 2020. Our findings indicate that there is a high correlation between the consumer confidence indicator based on survey data, and the consumer sentiment that we compute using data from Twitter platform. With our approach, we are even able to forecast the change in next month's consumer confidence.

Keywords: consumer sentiment; consumer confidence; Twitter; sentiment analysis

*Financial support by the Nordakademie Foundation through the project "Neuvermessung der Weltwirtschaft" is gratefully acknowledged.

[†]Corresponding Author: Department of Economics, Universität Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany, Phone: +49-40-42838-8151, Fax: +49-40-42838-8259, E-mail: deniz.oersal@uni-hamburg.de

[‡]E-mail: silke.sturm@uni-hamburg.de

1 Introduction

The main aim of this study is to propose a new consumer confidence indicator based on social media data for Germany, which may help us to forecast the turning points in the economy. To achieve this goal we are applying the sentiment analysis and proposing a new sentiment score. The existing sentiment scores in the literature use the polarity (negative, positive and neutral) of the words. In other words, the percentage of negative and positive words in a document (e.g. sentence, paragraph, topic) are only considered to compute the sentiment score. Different from these sentiment scores, which check the polarity of the words, we base our sentiment score on the values of the words which are relevant for the sentiment analysis. The sentiment score that we propose delivers satisfactory results.

Many scholars in the literature highlight that consumer sentiment is a good indicator for forecasting household spending (cp. Mueller, 1963; Carroll et al., 1994; Vuchelen, 2004; Ludvigson, 2004). Especially, according to Vuchelen "Consumer sentiment is meticulously watched during recessions since any significant change or lack thereof, is considered a very valuable sign of a nearby turning point or prolongation of the trough." (Vuchelen, 2004, p.494). Therefore, consumer confidence is an important leading indicator in macro-economics, which may help to find out how the private consumption will change. In other words, the sentiment score that we develop could be used to predict private consumption.

In this study we use the data from the social media platform Twitter to compute our social media based consumer confidence indicator. Lansdall-Welfare et al. (2012) indicate that *Tweets* could be used to nowcast the mood of the nation. Therefore, in this study we will use Twitter data to nowcast/forecast the consumer sentiment in Germany. With this prediction we believe that we may forecast the consumer confidence in Germany before the official survey based statistics are announced.

This paper is organized as follows: Section 2 will give a short literature review on forecasting important aggregate financial and macro-economic indicators via Twitter data. Section 3 will describe the data and the methodology. Section 4 will summarize the main findings of the quantitative analysis and Section 5 will conclude. The figures are deferred to the Appendix.

2 Literature Review

In the last decade the scholars in economics benefited from the availability of extensive and huge online text data (e.g. social media messages, blogs, forums, newspaper articles) and used them to forecast for the economy relevant indicators. For example, Twitter data has been used among others to predict the leading stock market indices. Zhang et al. (2011) used mood words such as hope, happy, fear, worry, nervous, anxious, upset to forecast the stock market indices, such as Dow Jones, NASDAQ and S&P 500. Using a randomized sample of tweets for a period of six months (30. March 2009 until 7. September 2009), their analysis showed that the mood on Twitter can help to predict the mood on the next day in stock exchange markets. In their study Bollen et al. (2011) also found that the value of

the daily Dow Jones Industrial Average (DJIA) is highly correlated with the public mood states derived from Twitter feeds. With a Granger Causality analysis they have also demonstrated that the public mood states measured via Twitter feeds can help to predict the daily up and downs in DJIA. Furthermore, Tweets have been also used to forecast box office revenues (Asur and Huberman, 2010).

In addition to these applications, data from Twitter platform have been also used to predict consumer confidence. Dass and Puts (2014) used messages from the Dutch public social media (from Twitter, Facebook, blogs, forums etc.) to forecast the changes in consumer confidence with changes in social media sentiment. Their analysis revealed that the Dutch survey based consumer confidence indicator (CCI) and the social media sentiment are highly correlated, and that even a long-run equilibrium relationship between them (i.e. cointegration relationship) is present. In another study, Shayaa et al., 2018 used public social media data (mostly Twitter data) to estimate the correlation between the official consumer confidence indicator and the sentiment from social media in Malaysia. They used keywords such as "car" and "holiday", and collected public data for the period from January 2015 to December 2016. However, their results demonstrate a significant but a weak association between the official CCI and the sentiment from social media. An extensive literature review on the application of social media data for predicting decisions of consumers can be found in Karaman Örsal (2021).

3 Data

There are several survey-based consumer confidence indicators in Germany. In this study we consider six different consumer confidence indicators: (1) Consumer confidence indicator (CCI) of Organization for Economic Cooperation and Development (OECD), (2) future expectations of the consumers in Germany reported by OECD.Stats, (3) ICON Consumer Confidence Indicator, (4) Consumer confidence indicator based on consumer survey of European Commission (EC), (5) IPSOS Consumer Confidence Indicator of Thomson Reuters and (6) Consumer Barometer of "Handelsverband Deutschland" (HDE). For example, OECD publishes monthly CCI, in which the value 100 represents the normal situation, value above 100 shows that the consumers are optimistic about the general economic situation and the financial household conditions. On the contrary, a value below 100 indicates that the consumers are pessimistic about their financial situation, unemployment expectations and the general economic situation¹. The respondents answer questions related to the general economic situation, financial situation, price trends in the last 12 months and in the next 12 months, unemployment expectations, major purchases, savings at present and in the next 12 months.

The online platform Twitter offers a micro-blogging service. Since 2017 due to a change in Twitter's terms of use each tweet has up to 280 characters. Every tweet has a unique identification (ID) number. Using the Application Programming Interface (API) of Twitter, we have randomly collected public tweets from Germany.

¹http://www.oecd.org/sdd/leading-indicators/glossaryforoecdcompositeleadingindicators.htm#CONSUMER_CONFIDENCE_INDICATOR_Standardised

Up to now we have downloaded tweets from 01. May 2019 until 31. July 2020. To scrap the public tweets we used the keywords "*Wirtschaft*" (Economy), "*Kaufen*" (Buying) and "*Kauf*" (Buy) which are the most related keywords with the questions in the consumer confidence surveys.

Table 1: Number of Tweets

Month	Kaufen		Wirtschaft		Kauf	
	Raw	w/o dup. ID	Raw	w/o dup. ID	Raw	w/o dup. ID
May 2019	12,125	11,567	12,608	12,464	320	220
June 2019	20,693	20,172	14,087	13,806	378	277
July 2019	20,311	19,790	15,170	14,907	353	253
Aug 2019	20,427	19,907	14,956	14,643	333	233
Sep 2019	19,633	19,133	17,566	17,066	452	252
Oct 2019	18,846	18,306	16,595	16,055	433	241
Nov 2019	20,360	19,860	16,228	15,728	743	505
Dec 2019	19,802	19,278	15,878	15,352	5,947	5,430
Jan 2020	19,879	19,339	19,121	18,581	7,096	5,835
Feb 2020	19,885	19,331	16,904	16,369	6,681	6,181
March 2020	31,291	30,776	33,118	32,598	8,035	7,515
April 2020	25,583	25,066	35,768	35,247	8,393	7,873
May 2020	22,604	22,084	27,071	26,551	7,434	6,914
June 2020	18,693	18,125	18,026	17,459	4,565	4,243
July 2020	22,885	21,140	18,793	17,312	8,256	6,633
Sum	313,017	303,874	291,889	284,138	5,9419	5,2605
All 3 Keywords	664,325	640,617				
Kauf+Kaufen	372,436	356,479				

As Table 1 shows, we initially had in total 664,325 *Tweets* (and 372,436 *Tweets* with the keywords "Kauf" and "Kaufen") for the analysis. After *Tweets* with the same ID number were removed, in total 640,617 messages remained for the sentiment analysis. Depending on the keywords, we used an average of 23,765 to 42,707 public tweets per month from Germany. This number is significantly higher than the number of respondents (e.g. 2000) used for the consumer surveys conducted each month.

3.1 Sentiment Analysis

As a text mining tool, sentiment analysis is used to find out opinions, feelings, responses and judgments (c.f. within a text Dey et al., 2019). The sentiment within a text could be either negative, positive or neutral. Lately, it became more important to look at the sentiments in order to predict human behavior. Therefore, sentiment analysis is the most appropriate tool for our purposes.

In this study, the consumer sentiment from Twitter data will be computed by using a dictionary-based sentiment analysis. Before the computation of the sentiment score we have removed the usernames, uniform resource locators (url), numbers,

punctuation and converted all tweets to lower case. Afterwards we grouped all the tweets from the same month.

The sentiment score is computed by using the sentiment value of each word based on the dictionary approach. Thereby, we use the sentiment dictionary SentiWS for the German language, which was developed by Remus et al. (2010). In this dictionary adverbs, adjectives and words which are relevant for the sentiment analysis take values between -1 and 1 according to their negative or positive sentiments. The monthly sentiment score formula is then:

$$Score_d = \frac{\sum_{w \in q} f(w, d) v_w}{N_d}, \quad d = \text{May 2019}, \dots, \text{July 2020}, \quad (1)$$

where d is the document and here the document represents each month, v_w is the sentiment value for each word, $f(w, d)$ is the frequency of word w in document d , q is the dictionary, and N_d is the number of total words in each month.

4 Results

Within the analysis three different social media-based sentiment scores are computed. The first sentiment score ($Score_d^{wirt-kauf}$) is computed with *Tweets* scraped using all the three keywords (i.e. "Wirtschaft", "Kauf", "Kaufen"), the second sentiment score ($Score_d^{wirt}$) is calculated with the *Tweets* scraped using the word "Wirtschaft" (i.e. "economy") and the last sentiment score (i.e. $Score_d^{kauf}$) is computed using the words "Kauf" and "Kaufen" (i.e. "to buy" and "buying"). For a better visualisation of the results both the survey-based consumer confidence indicators and the social media-based sentiment scores are converted to z-scores.

As it is apparent from Figures (1) - (6), with the social media based sentiment scores we may predict the direction of the change in consumer confidence one-month in advance. This is also verified with the highly significant positive correlation coefficients between the one-month lagged social media sentiment scores and the survey-based consumer confidence indicators, which lie within the range [0.829, 0.940] (see Table 2), especially with the sentiment score computed with the *Tweets* scraped using all the three keywords (i.e. $Score_d^{wirt-kauf}$). These figures are higher than the correlation coefficient with the social media sentiment scores in levels (c.f. $\rho = 0.878$ and 0.474). The lowest correlation coefficients are observed between the HDE consumer barometer and the sentiment scores in levels. Furthermore, the correlation coefficients of the other two sentiment scores are lower than the sentiment score based on all the three keywords used to scrap the *Tweets* (i.e. $Score_d^{wirt-kauf}$). Nevertheless, according to the Figures (1) - (6), the sentiment score based on the *Tweets* related with "to buy" and "buying" reacts much earlier to the changes in the sentiment than the sentiment score based on the *Tweets* related with the keyword "economy".

Table 2: Correlation Matrix

Indicator	Social Sentiment			Lagged Social Sentiment		
	$Score_d^{wirt-kauf}$	$Score_d^{wirt}$	$Score_d^{kauf}$	$Score_{d-1}^{wirt-kauf}$	$Score_{d-1}^{wirt}$	$Score_{d-1}^{kauf}$
CCI_EC	0.8128 (0.0002)	0.7839 (0.0005)	0.5210 (0.0460)	0.9200 (0.0000)	0.7660 (0.0180)	0.7590 (0.0220)
EXPECT	0.8775 (0.0000)	0.8459 (0.0001)	0.4910 (0.0630)	0.8500 (0.0010)	0.6890 (0.0950)	0.7200 (0.0520)
HDE	0.4740 (0.0743)	0.3949 (0.1451)	0.3840 (0.1570)	0.8290 (0.0030)	0.7350 (0.0380)	0.5120 (1.000)
ICON	0.8189 (0.0002)	0.7813 (0.0006)	0.5360 (0.0400)	0.9400 (0.0000)	0.8110 (0.0050)	0.7760 (0.0140)
IPSOS_CCI	0.7203 (0.0025)	0.6668 (0.0066)	0.4850 (0.0670)	0.9370 (0.000)	0.8330 (0.0030)	0.7490 (0.0270)
CCI_OECD	0.8165 (0.0002)	0.7323 (0.0019)	0.6380 (0.0110)	0.9090 (0.0000)	0.7110 (0.0630)	0.8250 (0.0030)

Notes: Bold figures show the correlation coefficients which are significant at the 5% significance level. $Score_d^{wirt-kauf}$ denotes the social sentiment score computed based on tweets scraped using the keywords “Wirtschaft”, “Kaufen” and “Kauf”, $Score_d^{wirt}$ denotes the social sentiment score computed based on *Tweets* scraped using the keyword “Wirtschaft”, $Score_d^{kauf}$ denotes the social sentiment score computed based on tweets scraped using the keywords “Kaufen” and “Kauf”, for month d .

5 Conclusions and Discussion

This study shows that the monthly consumer sentiment could be nowcasted or even forecasted one month ahead by using Twitter data. The results point out a highly significant positive association between the official survey-based consumer confidence statistics especially from OECD, IPSOS, ICON and EC, and the social sentiment measured with Twitter data.

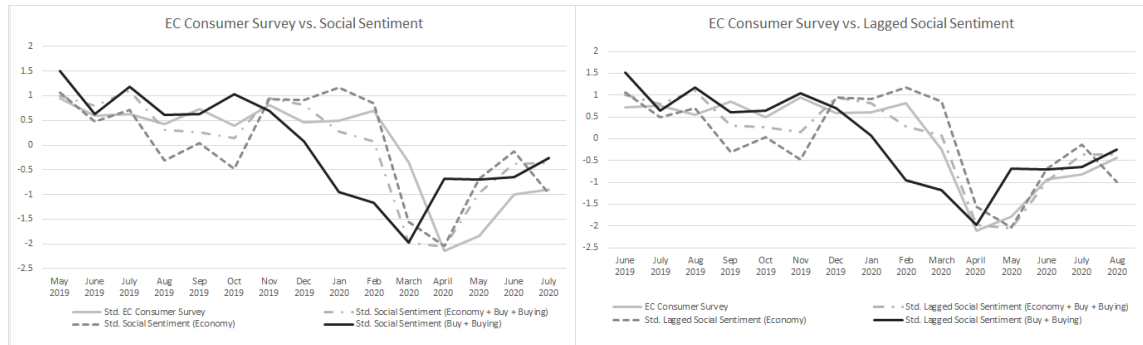
In other words, via consumer sentiments based on social media, it may be even possible to forecast the turning points of the German economy and to avoid time consuming and expensive consumer surveys.

References

- Asur, S., Huberman, B.A., 2010. Predicting the Future With Social Media. Technical Report. URL: <http://arxiv.org/abs/1003.5699>.
- Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2, 1–8.
- Carroll, C.D., Fuhrer, J.C., Wilcox, D., 1994. Does consumer sentiment forecast household spending? If so, why. *The American Economic Review* 84, 1397–1408.
- Dass, P.J.H., Puts, M.J.H., 2014. Social Media and Consumer Confidence. Technical Report. Workshop on Using Big Data for Forecasting and Statistics.

- Dey, N., Borah, S., Babo, R., Ashour, A.S., 2019. Social network analytics:: Computational research methods and techniques.
- Karaman Örsal, D.D., 2021. Online-Daten und Konsumententscheidungen, in: Straubhaar, T. (Ed.), Neuvermessung der Datenökonomie. Hamburg University Press.
- Lansdall-Welfare, T., Lampos, V., Nello, C., 2012. Nowcasting the mood of the nation. *Significance* 9, 26–28.
- Ludvigson, S.C., 2004. Consumer confidence and consumer spending. *Journal of Economic Perspectives* 18, 29–50.
- Mueller, E., 1963. Ten years of consumer attitude surveys: Their forecasting record. *Journal of the American Statistical Association* 58, 899–917.
- Remus, R., Quasthoff, U., Heyer, G., 2010. Sentiws - a publicly available german-language resource for sentiment analysis, in: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta.
- Shayaa, S., Ainin, S., Jaafar, N.I., Zakaria, S., Phoong, S.W., Yeong, W.C., Al-Garadi, M.A., Muhammad, A., Piprani, A.Z., 2018. Linking consumer confidence index and social media sentiment analysis. *Cogent Business and Management* 5, 1–12. doi:DOI: 10.1080/23311975.2018.1509424.
- Vuchelen, J., 2004. Consumer sentiment and macro-economic forecasts. *Journal of Economic Psychology* 25, 493–506.
- Zhang, X., Fuehres, H., Gloor, P.A., 2011. Predicting stock market indicators through twitter "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences* , 55–62.

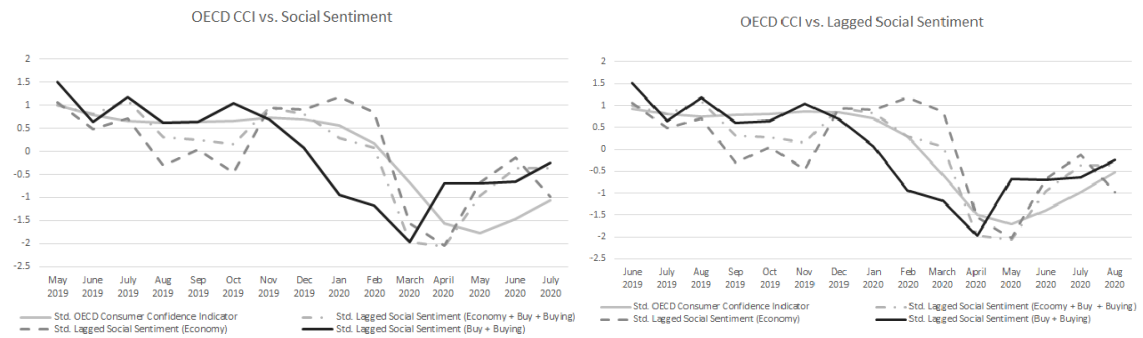
Appendix



(a) Sentiment Score and EC Consumer Survey

(b) Lagged Sentiment Score and EC Consumer Survey

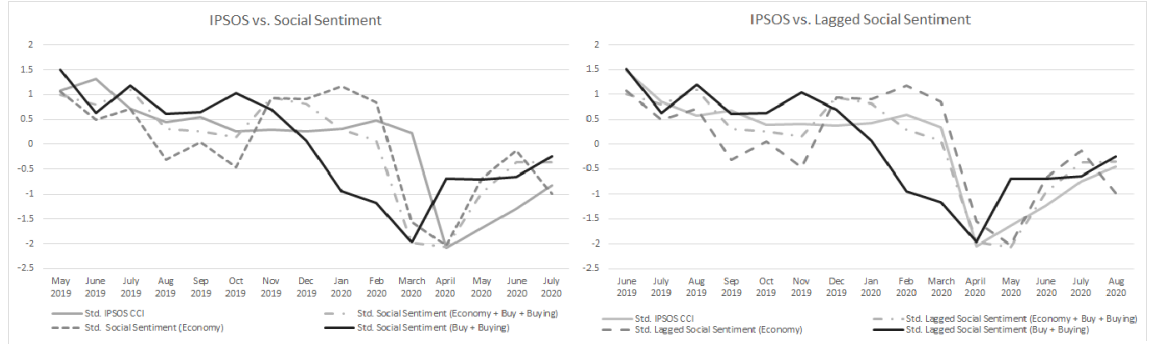
Figure 1: EC Consumer Survey and Social Sentiment Score from Twitter



(a) Sentiment Score and CCI of OECD

(b) Lagged Sentiment Score and CCI of OECD

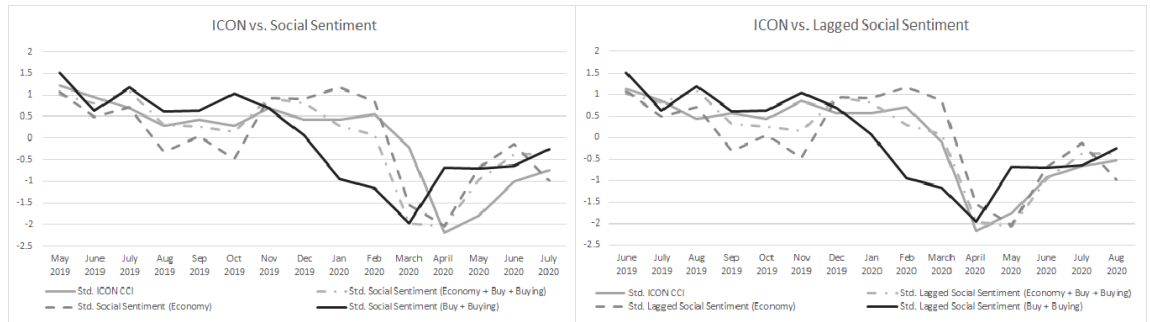
Figure 2: OECD Consumer Confidence Indicator and the Social Sentiment Score from Twitter



(a) Sentiment Score and IPSOS CCI

(b) Lagged Sentiment Score and IPSOS CCI

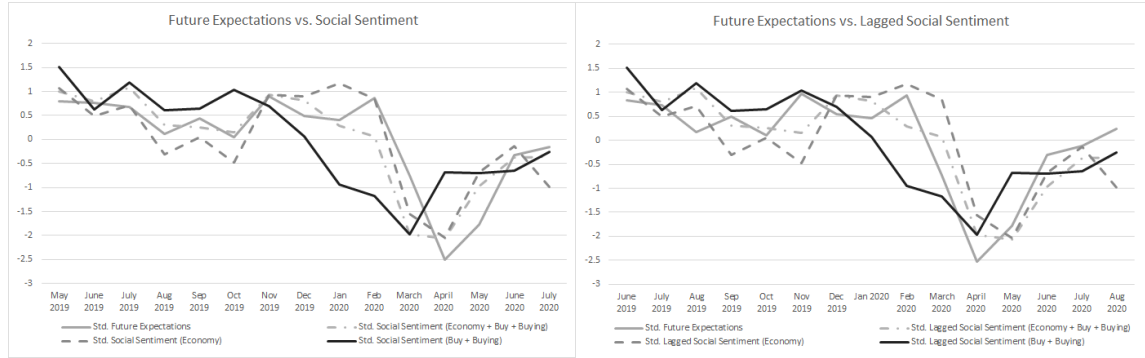
Figure 3: IPSOS Consumer Confidence Indicator and the Social Sentiment Score from Twitter



(a) Sentiment Score and ICON CCI

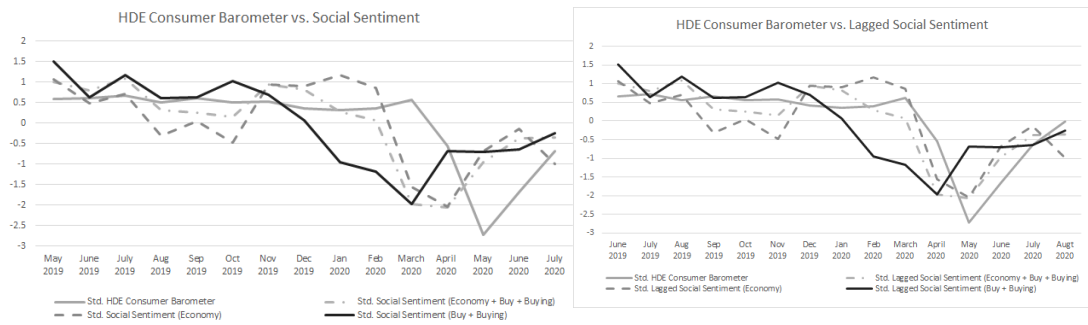
(b) Lagged Sentiment Score and ICON CCI

Figure 4: ICON Consumer Confidence Indicator and the Social Sentiment Score from Twitter



(a) Sentiment Score and Expectations for Future (b) Lagged Sentiment Score and Expectations for Future

Figure 5: Expectations for Future and the Social Sentiment Score from Twitter



(a) Sentiment Score and HDE Consumer Barometer (b) Lagged Sentiment Score and HDE Consumer Barometer

Figure 6: HDE Consumer Barometer the Social Sentiment Score from Twitter