Shayaa et al., *Cogent Business & Management* (2018), 5: 1509424
https://doi.org/10.1080/23311975.2018.1509424

cogent business & management

MARKETING | RESEARCH ARTICLE

Linking consumer confidence index and social media sentiment analysis

Shahid Shayaa[1], Sulaiman Ainin[2*], Noor Ismawati Jaafar[2], Shamsul Bahri Zakaria[2], Seuk Wai Phoong[2], Wai Chung Yeong[2], Mohammed Ali Al-Garadi[3], Ashraf Muhammad[4] and Arsalan Zahid Piprani[2]

Received: 20 June 2018
Accepted: 03 August 2018
First Published: 10 August 2018

*Corresponding author: Sulaiman Ainin, Department of Operations and MIS, Faculty of Business and Accountancy, University of Malaya, Kuala Lumpur, Malaysia
E-mail: ainins@um.edu.my

Reviewing editor:
Len Tiu Wright, De Montfort University Faculty of Business and Law, UK

Additional information is available at the end of the article

Abstract: This study aims to analyse the link (correlation) between and the official CCI and social media big data (via sentiment analysis) on consumer purchasing behaviour for two types of products over the course of two years (24 months, from January 2015 to December 2016). The CCI data was obtained from the Malaysian Institute of Economic Research (MIER) while the sentiment analysis was obtained from twitter. The results indicate that there is a significant but very small relationship between CCI and social media sentiment analysis. On the basis of the results we conclude that social media can offer huge a huge volume of data on consumer confidence, the analysis of which can be conducted at a more rapid time and integrated with existing methods in a synergistic way to refine the accuracy of the CCI using data from far larger populations.

Subjects: Consumers; Data Analytics; Social Media

Keywords: sentiment analysis; consumer confidence index; correlations

1. Introduction

An accurate gauge of consumer confidence is important in several respects. Consumer confidence gives insights into how positive/negative consumers feel regarding their personal financial

ABOUT THE AUTHOR

The authors for this paper consists of three groups: academics, industry and students. The academics are Ainin Sulaiman (principal researcher), Noor Ismawati Jaafar, Shamsul Bahri Zakaria, Phoong Seuk Wai and Yeong Wai Chung. They are all from the Department of Operations and MIS, Faculty of Business and Accountancy, University of Malaya. Shahid is from Berkshire Media Sdn Bhd, a fast growing social analytics firm while Arslan and Dr AL-GARADI (now in University of Qatar) are the research assistants and Dr Ashraf (now in COMSATS University Islamabad) was the post-doctoral student.

PUBLIC INTEREST STATEMENT

This study examines consumer confidence using two types of data: consumer confidence index (CCI) and social media data via sentiment analysis. CCI is based on a household survey conducted quarterly. Respondents are asked to evaluate their household's current and expected financial positions, employment outlook and purchasing behaviour. Sentiment analysis on the other hand is the process of determining whether a text is positive, negative or neutral. These text (related to purchasing behaviour) are extracted from social media, such as Twitter and analysed using tools, such as Machine language. The results indicate that there is a significant but very small relationship between CCI and sentiment analysis. Hence, it is concluded that sentiment analysis can offer a huge volume of data on consumer confidence, the analysis of which can be conducted faster hence businesses would be able to predict consumer confidence and take the necessary strategies to improve their sales.

capability and purchasing behaviour, and regarding the economy as a whole. With a clear understanding of consumer confidence, organisations will be able to change their strategy to align with the current consumer environment. For example, if consumer confidence is found to be low, organisations can focus on products/services to more value conscious consumers, and to align marketing campaigns according to the current consumer confidence environment. It is also vital knowledge for governments when formulating plans for reviving the economy, enhancing consumers' confidence and encouraging them to spend. Considering the importance of consumer confidence, an accurate tool for gauging consumer confidence by means of a measureable value is greatly needed.

Consumer confidence surveys are regularly conducted in at least 45 countries (Curtin, 2007). Most countries have their own indices that measure consumer confidence. These indices are based on household surveys. In Malaysia, the Malaysian Institute of Economic Research (MIER) publishes the Consumer Confidence Index (CCI), which gives a numerical value for consumer confidence. The CCI is based on surveys conducted quarterly on a sample of over 1,200 households. Respondents are asked to evaluate their household's current and expected financial positions, employment outlook and purchasing behaviour. Questions relating to plans to buy houses, new or used cars and other major consumer durables are also asked.

Relying on the CCI as an economic indicator commonly raises two dilemmas (Ludvigson, 2004): whether to focus on index-level or month-to-month changes, and whether to focus on the present conditions or the expectation component. Ludvigson (2004) raised concerns about the effectiveness of the survey-based CCI by that it would hardly to find that confidence surveys reflect the current state of consumer purchasing behaviour.

On the other hand, the social media which is built on Web 2.0 technologies brings a new way for the general public to show their attitudes towards products and services in a form that may ultimately influence other prospective consumers (Rashidi, Abbasi, Maghrebi, Hasan, & Waller, 2017). Social media is now ingrained into human lives and integrated with the core constituents of various human activities and modes of communication (Lassen, Madsen, & Vatrapu, 2014). Social media users frequently express their opinions and views on a wide range of topics, including products and services (Arora, Li, & Neville, 2015). For example, users might express their confidence (or lack of confidence) in the state of the economy, disappointment over certain products or services, or express their opinions, emotions, feelings and problems regarding their past purchases or future purchase intentions, etc. In businesses, social media platforms (e.g. Facebook, Twitter, blogs and forums) are now used to connect to existing customers, market products and brands, and explore new business opportunities. They enable the organisation to learn about shoppers' experiences and understand consumers' buying behaviours, which in turn facilitates improvements in their marketing and customer services efforts.

In view of the information that can be obtained from social media, a number of studies have been done on sentiment analysis of social media. Various techniques for opinion mining, text analytics and sentiment analysis have also been documented and reported for understanding various aspects of social media data. Sentiment analysis and opinion mining are fields of study that analyse people's opinions, evaluations, attitudes and emotions, usually from written language (Liu, 2012). There are many ways to analyse sentiments, including machine-learning, lexicon-based, keyword-based, and concept-based approaches (Cambria, 2016). Most studies of social media sentiments are based on textual information. Piryani, Madhavi, and Singh (2017) recently provided an analytical mapping of studies done on opinion mining and sentiment analysis between 2000 and 2015.

Although a number of studies have been undertaken on consumer sentiments through social media, less attention has been given to studying the relationship between the consumer confidence predictors of purchasing behaviour derived from social media and from the official CCI.

Hence, this study would like to examine the relationship between the sentiments expressed through social media big data and the official CCI produced by MIER. More specifically, this study aims to evaluate the correlation between social media big data (derived via sentiment analysis) on consumer purchasing behaviour for two types of products and the equivalent data from the official CCI, between January 2015 and December 2016.

2. Literature review

2.1. Consumer confidence index

The CCI was developed as a source of economic statistics during the mid-20th century, and has become a barometer whose results affect economic policy, stock markets and broader government policy. The idea of a CCI for forecasting consumer spending was first proposed by Mueller (1963). She found that consumer confidence was a significant explanatory variable in consumption spending. Several studies have been conducted to examine the influence of consumer confidence. Mehra and Martin (2003) found that consumer confidence was a significant predictor in regression equations for consumer spending. Garrett, Hernández-Murillo, and Owyang (2005) used regional data to show that consumer confidence helps to predict retail spending in U.S. states. Vuchelen (2004) studied the relationship between consumer sentiment, expected income and the uncertainty surrounding this income. In terms of household spending, Acemoglu and Scott (1994) and Carroll, Fuhrer, and Wilcox (1994) have found that lagging consumer sentiment has significant explanatory power for current changes in household spending in the United States and Britain. Another study on consumer sentiment by Yacob and Mahdzan (2014) found that consumer confidence volatility has significant predictive power for stock market volatility.

In contrast, Ludvigson (2004) raised the question: do consumer confidence surveys really provide information that predicts the future path of household spending? The results indicated that survey measures do contain some information about the future path of consumer spending but fall short of capturing real consumer purchasing behaviours. It would hardly to find that confidence surveys reflect the current state of consumer purchasing behaviour and the overall economy. That is because of the two dilemmas that CCI analyses have commonly encountered: whether to focus on index-level or month-to-month changes, and whether to focus on the present conditions or the expectation component. However, a conjecture now arises that the confidence indexes might prove to be even more useful if based on sentiment analysis of consumers' social media data rather than on consumer surveys. Consequently, the question to be answered is: would indexes of consumer confidence based on survey data or on social-media big-data sentiment analysis be more valuable for forecasting consumer purchasing behaviour in real time? In other words, should a researcher/policy analyst use data reported from consumer confidence surveys or consumer social-media sentiment to improve forecasts of consumer purchasing behaviour? This issue has been examined in the literature previously, with the conclusion that confidence-survey data helped improve the forecasts slightly, but not statistically significantly. For example, Croushore (2005) found that the confidence survey indexes are not of significant value in forecasting consumer spending. In fact, in some cases, they make the forecasts significantly worse, suggesting that consumer-confidence surveys are no better than social-media sentiment data in capturing information about consumer purchasing behaviour. However, we conjecture that using consumer sentiment data on social media might show greater marginal significance for consumer confidence, because the surveys might capture effects that will not appear in the data.

However, there are several weaknesses with the method currently adopted to obtain the CCI data. First, it is time consuming and costly to conduct surveys with a large sample. Besides which, the households chosen may not be representative of the whole population because there are limitations on the number of respondents that can be chosen due to cost and time constraints. Second, the surveys are conducted quarterly, and there is only a single CCI value for each quarter. Hence changes in consumer confidence within a certain quarter, for example between different months in the same quarter, cannot be captured by the CCI value. Third, not all respondents will complete the surveys. Fourth, the CCI is only based on responses to the questions that are asked during the survey, and other aspects of consumer confidence may not be captured in the CCI. Hence, the official CCI is assumed to reflect nothing else but the answers to the survey questions.

2.2. Social media sentiment analysis

The digital age has accelerated the growth of social media networks (Rashidi et al., 2017). It has also created new opportunities for the consumers to express their opinions and reviews of products and services (Arora et al., 2015). As a result of this, enormous data resources have been created over the years, and how to capitalise on these sources of data turns out to be a striking topic for researchers from various disciplines, including computer science, social sciences, economics, mathematics and management. Scholars and practitioners from these various fields have devoted their energies to obtaining meaningful information from these data sources. Various techniques in the forms of opinion mining, text analytics and sentiment analysis have been documented and reported for analysing various aspects of social media big data. In recent times, numerous studies have been conducted on utilising user-generated data on social media. Daas and Puts (2014) have shown that there is a clear association between changes in the sentiment of social media messages and consumer confidence. On the other hand, Dong and Cooper (2016) used sentiment analysis for product recommendations, and D'Avanzo and Pilato (2015) utilised Facebook to mine users' opinions for assistance with shopping decisions. Sentiment analysis has also been used in the field of finance. For example, Guo, Sun, and Qian (2017) exploited Chinese investor sentiment for predicting stock prices. Rao and Srivastava (2014) employed Twitter sentiments to forecast the US stock market. However, changes in consumer sentiment not only predict the changes in consumer spending, but also cause them. An alternative interpretation could be that sentiment analysis of social media big data more accurately forecasts consumer purchasing behaviour than survey data because it reflects the true overall outlook for the economy.

The massive popularity of social media channels like Facebook, Twitter, LinkedIn, blogs and forums has spawned a massive amount of data. According to the International Data Corporation (IDC), the amount of data generated up to 2010 was 1 ZB, and data production has accelerated explosively since then, generating 7 ZB by the end of 2014. Twitter now generates 175 million tweets on a daily basis and has more than 467 million users (Arora et al., 2015). This rapidly expanding usage of social media platforms drives the application of social media data analytics (Thackeray, Neiger, Hanson, & McKenzie, 2008). Opinions that are mined through social media data analytics can be used to gain insights about consumers' sentiments towards any product or service. Firms can capitalise by analysing this rich data to obtain valuable insights and hidden knowledge, giving them to acquire competitive edge (He, Zha, & Li, 2013). The intelligence obtained through analysing sentiments can be the driver for the businesses to gain rich insights for making high-impact decisions.

Various techniques, such as opinion mining, text analytics and sentiment analysis, have been documented and reported to analyse various aspects of social media big data. The social-media data-analytic technique of sentiment analysis could be utilised to measure consumer confidence effectively (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011). In recent times, numerous studies have been conducted on the utilisation of user-generated data from social media. Daas and Puts (2014) have shown that there is a clear association between changes in the sentiment of social media messages and consumer confidence. On the other hand, Dong and Cooper (2016) used sentiment analysis for product recommendations, and D'Avanzo and Pilato (2015) applied a multi-facet sentiment analysis technique to predict the sales of mobile applications. D'Avanzo and Pilato (2015) utilised Facebook to mine users' opinions for assistance with shopping decisions. Zhou, Xia, and Zhang (2016) applied a multi-granularity approach to investigate online shopping behaviour. Sentiment analysis has also been used in the field of finance. For example, Guo et al. (2017) exploited Chinese investor sentiment to predict stock prices. Bollen, Mao, and Zeng (2011) and Rao and Srivastava (2014) employed Twitter sentiments to forecast the US stock market.

Unlike the official CCI data, which is obtained through household surveys where the number of respondents is constrained by time and cost considerations, millions of social media messages can be analysed easily. Moreover, a daily, weekly or monthly numerical measure can be created through the analysis of social media messages, as a high volume of messages are uploaded daily. Therefore, changes in consumer confidence can be captured within a shorter time frame than household surveys could measure. Users also tend to be more truthful when they are expressing their opinions through social media than when they are filling out a survey. Lastly, whereas surveys can only elicit information in the form of responses to a list of predetermined questions, social media users can talk about basically anything; hence more information from a variety of aspects can be obtained.

3. Research methodology

In this section, we describe how the data is extracted, how sentiment analysis is applied, how the CCI data is constructed then how correlation between social media sentiment analysis and CCI is measured.

3.1. Data collection

The main aim of the study is to analyse the big data on consumer sentiment from social media, then investigate its correlation with the CCI. Consequently in this study, the big data was extracted from social media platforms. Berkshire Media Sdn Bhd, a social analytics firm, extracted posts from Twitter, forums, mainstream media, blogs, Facebook, online comments and YouTube. Most of the data for this paper were extracted from Twitter. The Twitter application program interface (API) enables the researcher to extract public posts easily (Pak & Paroubek, 2010). Twitter has been used as source of data for many researchers compared to other social media websites (Al-Garadi, Varathan, & Ravana, 2016). Only data that are made public are extracted. The extracted data are from January 2015 to December 2016, which represents a huge volume of user-generated opinion.

3.2. Sentiment analysis

The Berkshire Media Sdn Bhd social analytics firm has developed a supervised machine-learning-based classifier. First, this machine analyses multilingual posts and classifies them into positive, negative and neutral. Each post is then given a sentiment score based on how positive or negative the posts are. Positive posts are given a positive sentiment score, neutral posts are given a zero sentiment score, while negative posts are given negative sentiment scores. Second, the data extraction involves categorising the purchasing behaviour posts into two categories: car and holiday. The extraction was carried out using automatic categorisation based on the keywords in English and Bahasa Malaysian: 1 – Car (*KERETA, Kereta, kereta, kreta Kerete, kerete, kete, Kete, keta, car*). 2 – Holiday (*travel, trvel, holiday, Holiday, cuti, Cuti*). A total of 19,900 posts were extracted, 17,900 related to cars and remaining 2,000 related to holiday.

Finally, we analyse each of the above-mentioned categories on a monthly basis. For example, for the holiday category we divide the sentiment data into 12 months for two years, generating 24 files for each month's holiday sentiment-analysis data. In same manner the other category is constructed, therefore dividing the sentiment data into 48 files (two categories * 12 months * 2 years), such that each file represents a month in one category and each category has sentiment data for 24 months (2015–2016). Based on these files, we generate two values (average sentiment scoring and average sentiment counts) for 24 months for each category as follows:

The average sentiment score and average sentiment counts are calculated for every month of each categories. Sentiment score and average sentiment counts are calculated as following:

$$\text{Average sentiment score} = \frac{\sum postive\ scores - \sum negative\ scores}{\sum number\ of\ postive\ and\ negative\ posts} \tag{1}$$

Where average sentiment scoring is based on (total score of each positive – negative of the posts) divided by total number of positive and negative posts

$$\text{And Average sentiment counts} = \frac{\sum postive\ counts - \sum negative\ counts}{\sum number\ of\ postive\ and\ negative\ posts} \quad (2)$$

Where the Average sentiment counts is based only on (number of positive – negative posts) divided by total number of positive and negative posts.

3.3. Consumer confidence index data

The CCI data was obtained from MIER, which releases the data every quarterly. Using the software MATLAB, monthly data was constructed. Hence, there are 24 values of CCI index, a value for each month for the two year period (2015–2016) to show monthly index.

3.4. Correlation between social media sentiment analysis and CCI

After constructing CCI values for 24 months, and average sentiment score and average sentiment counts for the same time period, it was then possible to analyse correlations between CCI and sentiment from social media big data for 2015 and 2016.

4. Results

In this section, we will firstly show the sentiment scores, average sentiment scores and counts (calculated as shown in Equations (1) and (2), respectively) for the 24 months in each category: (1) car 2) holiday (see Tables 1 and 2). The CCI data is shown in Table 3. Correlations between the CCI and the average sentiment scores and average sentiment counts are shown in Table 4 and Figures 1 and 2.

Table 1 shows average sentiment scores and average sentiment counts (calculated as shown in Equations (1) and (2), respectively) for buying a car in 2015 and 2016. Generally, the sentiment scores, average sentiment scores and counts were higher in 2016 compared to 2015. The highest average sentiment scores and counts for buying a car are in February 2016. The sentiment for 2016 maybe higher as consumer perceived the country's economy to be much better.

Table 2 shows the average sentiment scores and counts (calculated as shown in Equations (1) and (2), respectively) for going on holiday in 2015 and 2016. It is illustrated that the sentiment scores, average sentiment scores and counts for 2016 is higher than 2015. Consistently, for both years, the highest scores is Jan, November and December. This may be due to the school holiday's season that generally begins in November and ends in January.

Table 3 illustrates the CCI data for the two-year period (2015 and 2016). Generally it can be seen that the CCI for 2016 is much better than 2015. This is similar to the sentiment analysis data for both car and holiday. This may be because the Gross Domestic Product (GDP) and GDP per capita are higher in 2016 than 2015, thus reflecting the consumer confidence is higher (https://tradingeconomics.com/malaysia/gdp-per-capita). The highest CCI is in May 2016 while the lowest is in November 2015.

The correlations between the social media sentiment analysis scores and CCI data on purchase intentions for car and holiday were carried out and are illustrated in Table 4. The correlation results represented in Table 4 show that there is no significant correlation between sentiment analysis and CCI data on purchasing intentions in any of the two categories.

Figures 1 and 2 illustrate the relationship between CCI and the average sentiment scores for buying cars whereas Figure 2 illustrates the relationship between CCI and going for holiday.

Table 1. Sentiment scores and counts (Car 2015–2016)

Months	Sentiment scores	Positive counts (Pc)	Negative counts (NC)	(PC + NC)	Avg. sentiment score	Avg. sentiment counts
Jan	787	139	44	183	4.297814208	0.519125683
Feb	1145	213	61	274	4.177007299	0.554744526
March	1283	228	129	357	3.592436975	0.277310924
April	513	136	53	189	2.711640212	0.439153439
May	867	134	53	187	4.636363636	0.43315508
June	1076	173	98	271	3.968634686	0.276752768
July	1347	274	142	416	3.237980769	0.317307692
Aug	482	112	59	171	2.81871345	0.30994152
Sep	462	95	65	160	2.884375	0.1875
Oct	877	151	53	204	4.299019608	0.480392157
Nov	717	131	69	200	3.585	0.31
Dec	1999	244	77	321	6.22741433	0.520249221
Jan	1011	162	121	283	3.572438163	0.144876325
Feb	4247.5	492	56	548	7.750912409	0.795620438
Mar	1608.5	238	82	320	5.0265625	0.4875
Apr	610.5	113	80	193	3.163212435	0.170984456
May	1172.5	155	38	193	6.075129534	0.606217617
Jun	2028.5	254	58	312	6.501602564	0.628205128
Jul	800.5	126	63	189	4.235449735	0.333333333
Aug	788.5	122	55	177	4.45480226	0.378531073
Sept	901	158	88	246	3.662601626	0.284552846
Oct	857.5	135	67	202	4.245049505	0.336633663
Nov	745.5	111	50	161	4.630434783	0.378881988
Dec	619.5	97	46	143	4.332167832	0.356643357

5. Discussion

In this paper, the relationship was examined between the official CCI published by the Malaysian Institute of Economic Research and the sentiments expressed via social media in five purchasing-intention categories: (1) car, and 2) holiday. It was found that there is a correlation, but it is very small, almost negligible and non-significant. This result may be due to inherent differences between the official CCI and the consumer sentiments expressed on social media in terms of how data is collected and calculated. Due to the official CCI being calculated on the basis of survey data, there are several weaknesses with its data. For example, the households surveyed may not be representative of the whole population as there are limitations on the number of respondents that can be included due to cost and time constraints. Additionally, the CCI is only based on responses to the predetermined questions which are asked during the survey, meaning that other aspects of consumer confidence may not be captured. Unlike the official CCI, social media sentiment analysis is based directly on consumers' opinions, emotions, feelings and problems regarding their past purchases or future purchase intentions as these are expressed on social media. Users tend to be more truthful when they are expressing their opinions through social media, compared to when they filling out a survey. In addition, the social media platforms provide a bigger data set compared to the CCI, which is normally based on 1,000–2,000 respondents. However, it can be concluded that consumer confidence measured using social media data via sentiment analysis is not significantly correlated with the official CCI.

Table 2. Sentiment scores and counts (Holiday 2015)

Months	Sentiment scores	Positive counts (Pc)	Negative counts (NC)	(PC + NC)	Avg. sentiment score	Avg. sentiment counts
Jan	378.5	40	7	47	8.053191489	0.70212766
Feb	202	21	2	23	8.782608696	0.826086957
Mar	152	20	7	27	5.62962963	0.481481481
Apr	108.5	13	3	16	6.78125	0.625
May	125	15	4	19	6.578947368	0.578947368
Jun	191	19	8	27	7.074074074	0.407407407
Jul	119	19	8	27	4.407407407	0.407407407
Aug	167	16	0	16	10.4375	1
Sept	75.5	11	1	12	6.291666667	0.833333333
Oct	161.5	21	5	26	6.211538462	0.615384615
Nov	266.5	34	13	47	5.670212766	0.446808511
Dec	284	29	4	33	8.606060606	0.757575758
Jan	527.5	51	7	58	9.094828	0.758621
Feb	189	19	5	24	7.875	0.583333
Mar	175	29	9	38	4.605263	0.526316
Apr	111	14	3	17	6.529412	0.647059
May	125	15	4	19	6.578947	0.578947
Jun	473	47	8	55	8.6	0.709091
Jul	102.5	12	3	15	6.833333	0.6
Aug	97.5	12	4	16	6.09375	0.5
Sept	177	22	16	38	4.657895	0.157895
Oct	115	14	5	19	6.052632	0.473684
Nov	200	21	5	26	7.692308	0.615385
Dec	253.5	25	2	27	9.388889	0.851852

Table 3. CCI data for 2015 and 2016

Date	CCI month	Date	CCI month	Date	CCI month
Jan 2015	73.03628911	**Sept 2015**	67.56336547	**May 2016**	78.49015276
Feb 2015	72.59893704	**Oct 2015**	64.90118013	**Jun 2016**	77.51839204
Mar 2015	72.16158496	**Nov 2015**	63.83072419	**Jul 2016**	75.64225626
Apr 2015	71.82946628	**Dec 2015**	65.48565196	**Aug 2016**	73.60954417
May 2015	71.7078144	**Jan 2016**	68.97182138	**Sept 2016**	72.01057567
Jun 2015	71.77292506	**Feb 2016**	72.88814128	**Oct 2016**	70.80575544
Jul2015	71.48534338	**Mar 2016**	76.02918939	**Nov 2016**	69.79800935
Aug 2015	70.17667681	**Apr 2016**	77.97221912	**Dec 2016**	68.79026326

Table 4. Correlation between CCI and sentiment of big social media data for 2015 and 2016

Categories	*r* value (Avg. sentiment scores and CCI)	*r* value (Avg. sentiment counts and CCI)
Car	0.2473	0.24344
Holiday	−0.11009	−0.12947

Figure 1. Relationship between CCI and average sentiment scores: cars.

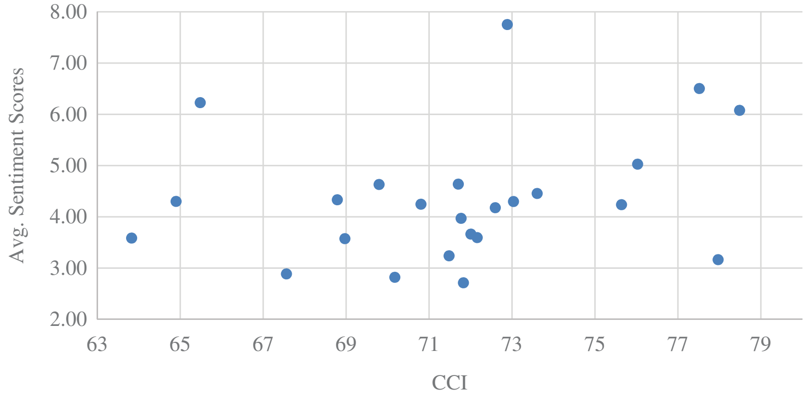
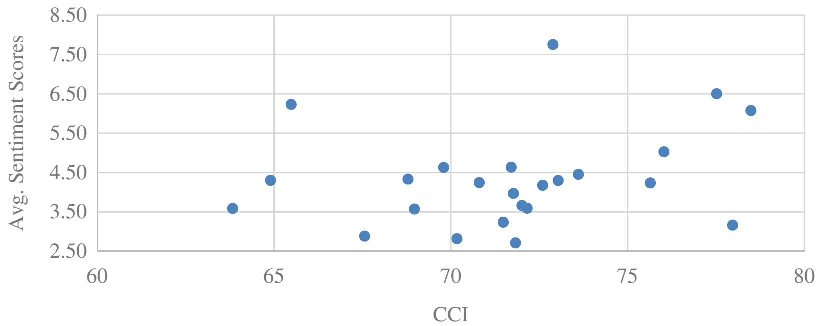


Figure 2. Relationship between CCI and average sentiment scores: holiday.



Now the question is, whether a researcher or policy analyst ought to rely on the official CCI or on social media sentiment as a measure of consumer confidence in order to improve forecasts of consumer purchasing behaviour? It can be explained and justified keeping in view on the past literature, with the conclusion that the official CCI helps to improve forecasts of consumer purchasing behaviour slightly, but not statistically significantly. Ludvigson (2004) raised concerns about the effectiveness of survey-based CCI by that it would hardly to find that confidence surveys reflect the current state of consumer purchasing behaviour and the overall economy. Additionally, Croushore (2005) found that the confidence survey indexes are not of significant value in forecasting consumer spending. In fact, in some cases they make the forecasts significantly worse, suggesting that consumer-confidence surveys are not better than social media sentiment data in capturing information about consumer purchasing behaviour. However, it can be concluded that using consumer sentiment data from social media might show greater marginal significance for measuring consumer confidence and thereby forecasting their purchasing behaviour, as it captures true information about consumer purchasing behaviour. Nevertheless, care needs to be taken with social media contents as these fluctuate in their positive and negative outcomes to their topics over time. The same individual may give positive sentiment at one point of time and perhaps in six months' time provide negative sentiment.

Conclusion

The aim of this study was to evaluate the relationship between the published CCI measures of Malaysian consumer purchasing behaviour (on car and holiday) and sentiment analysis of social media big data. It was found that there is a correlation, but that it is very small, almost negligible. Generally, it can be concluded that consumer confidence can also be measured using social media data via sentiment analysis. It can be captured almost instantly as posts are updated daily and sentiment scores can be calculated daily; hence changes in consumer confidence can be captured within a shorter time frame than survey methods make possible. In addition, users also tend to be more truthful when they are expressing their opinions through social media than when they are filling out a survey, which is the method used to calculate CCI. The household survey is based on responses toward a list of predetermined questions, whereas social media users can talk about basically anything; therefore, more information from a variety of aspects can be obtained. In addition, the social media data provides a bigger data set compared to the CCI, which is normally based on 1,000–2,000 respondents. In sum, future attempts to study consumer confidence should use social media big data. Future research must also include analysis that examines the reasons for the sentiment scores and CCI trends.

The methodology used in this study can be replicated in other countries. For example, the methodology used to determine the monthly sentiment analysis scores from social media data and CCI can be replicated to analyse consumer confidence in different countries. In addition, the average sentiment score and counts formulae can be adopted in future studies.

In conclusion, it can be said that organizations can use the social media sentiment analysis to predict their consumer purchasing behaviour and confidence as it has the advantage of enormous amount of data (volume), multiple sources of data (variety) and speed of data processing (velocity). The CCI data on the other hand does not have these capabilities but it has one important element i.e. its data is not uncertain and imprecise (veracity) thus making it more reliable.

Funding
This study is funded by the University of Malaya external grant (provided by Berkshire Media Sdn Bhd) [Grant number: PV003-2017].

Author details
Shahid Shayaa[1]
E-mail: shahid@berkshiremedia.com.my
Sulaiman Ainin[2]
E-mail: ainins@um.edu.my
ORCID ID: http://orcid.org/0000-0002-8989-712X
Noor Ismawati Jaafar[2]
E-mail: isma_jaafar@um.edu.my
ORCID ID: http://orcid.org/0000-0002-8604-6004
Shamsul Bahri Zakaria[2]
E-mail: esbi@um.edu.my
Seuk Wai Phoong[2]
E-mail: phoongsw@um.edu.my
Wai Chung Yeong[2]
E-mail: yeongwc@um.edu.my
Mohammed Ali Al-Garadi[3]
E-mail: mohammed.g@qu.edu.qa
Ashraf Muhammad[4]
E-mail: muhammadashraf@ciitvehari.edu.pk
Arsalan Zahid Piprani[2]
E-mail: arsalan@um.edu.my
[1] Berkshire Media Sdn Bhd, Petaling Jaya, Selangor, Malaysia.
[2] Department of Operations and MIS, Faculty of Business and Accountancy, University of Malaya, Kuala Lumpur, Malaysia.
[3] Research & Graduate Studies Office, University of Qatar, Qatar.
[4] Department of Management Sciences, COMSATS University Islamabad, Islamabad, Pakistan.

Citation information
Cite this article as: Linking consumer confidence index and social media sentiment analysis, Shahid Shayaa, Sulaiman Ainin, Noor Ismawati Jaafar, Shamsul Bahri Zakaria, Seuk Wai Phoong, Wai Chung Yeong, Mohammed Ali Al-Garadi, Ashraf Muhammad & Arsalan Zahid Piprani, *Cogent Business & Management* (2018), 5: 1509424.

References

Acemoglu, D., & Scott, A. (1994). Consumer confidence and rational expectations: Are agents' beliefs consistent with the theory? *The Economic Journal*, *104* (422), 1–19. doi:10.2307/2234671

Al-Garadi, M. A., Varathan, K. D., & Ravana, S. D. (2016). Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network. *Computers in Human Behavior*, *63* (October), 433–443. doi:10.1016/j.chb.2016.05.051

Arora, D., Li, K. F., & Neville, S. W. (2015). Consumers' sentiment analysis of popular phone brands and operating system preference using Twitter data: A feasibility study. In *Advanced Information Networking and Applications (AINA)*, September, *2015 IEEE 29th International Conference* (pp. 680–686). Adelaide, SA, Australia.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. doi:10.1016/j.jocs.2010.12.007

Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*, *31*(2), 102–107. doi:10.1109/MIS.2016.31

Carroll, C. D., Fuhrer, J. C., & Wilcox, D. W. (1994). Does consumer sentiment forecast household spending? If so, why? *The American Economic Review*, *84*(5), 1397–1408.

Croushore, D. (2005). Do consumer-confidence indexes help forecast consumer spending in real time? *The North American Journal of Economics and Finance*, *16* (3), 435–450. doi:10.1016/j.najef.2005.05.002

Curtin, R. (2007). Consumer sentiment surveys: Worldwide review and assessment. *OECD Journal. Journal of Business Cycle Measurement and Analysis*, *3*(1), 7–42. doi:10.1787/jbcma-v2007-art2-en

D'Avanzo, E., & Pilato, G. (2015). Mining social network users opinions' to aid buyers' shopping decisions. *Computers in Human Behavior*, *51*(PB), 1284–1294. doi:10.1016/j.chb.2014.11.081

Daas, P. J., & Puts, M. J. (2014). Social media sentiment and consumer confidence, *Paper for the workshop on using Big Data for forecasting and statistics*, Frankfurt, Germany, 7–8 April. Retrieved June 20, 2018, from http://www.ecb.europa.eu/events/pdf/conferences/140407/Daas_Puts_Sociale_media_cons_conf_Stat_Neth.pdf?409d61b733fc259971ee5beec7cedc61. Google Scholar.

Dong, Q., & Cooper, O. (2016). An orders-of-magnitude AHP supply chain risk assessment framework. *International Journal of Production Economics*, *182*, 144–156. doi:10.1016/j.ijpe.2016.08.021

Garrett, T. A., Hernández-Murillo, R., & Owyang, M. T. (2005). Does consumer sentiment predict regional consumption. *Federal Reserve Bank of St. Louis Review*, *87*, 123–135.

Guo, K., Sun, Y., & Qian, X. (2017). Can investor sentiment be used to predict the stock price? Dynamic analysis based on China stock market. *Physica A: Statistical Mechanics and Its Applications*, *469*, 390–396. doi:10.1016/j.physa.2016.11.114

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, *33*, 464–472. doi:10.1016/j.ijinfomgt.2013.01.001

Lassen, N. B., Madsen, R., & Vatrapu, R. (2014, September). Predicting iphone sales from iphone tweets. In *Enterprise Distributed Object Computing Conference (EDOC), 2014 IEEE 18th International* (pp., 81–90). Ulm, Germany.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures On Human Language Technologies*, *5*(1), 1–167. doi:10.2200/S00416ED1V01Y201204HLT016

Ludvigson, C. (2004). Consumer confidence and consumer spending. *The Journal of Economic Perspectives*, *18* (2), 29–50. doi:10.1257/0895330041371222

Mehra, Y. P., & Martin, E. W. (2003). Why does consumer sentiment predict household spending? *Economic Quarterly*, *89*(Fall), 51–67.

Mueller, E. (1963). Ten years of consumer attitude surveys: Their forecasting record. *Journal of the American Statistical Association*, *58*(304), 899–917. doi:10.1080/01621459.1963.10480677

Pak, A., & Paroubek, P. (2010, May 17–23). Twitter as a corpus for sentiment analysis and opinion mining. *Proceedings of the International Conference on Language Resources and Evaluation*, Valletta, Malta.

Piryani, R., Madhavi, D., & Singh, V. K. (2017). Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. *Information Processing & Management*, *53*, 122–150. doi:10.1016/j.ipm.2016.07.001

Rao, T., & Srivastava, S. (2014). Twitter sentiment analysis: How to hedge your bets in the stock markets. In F. Can, T. Ozyer, & F. Polat (Eds.), *State of the Art Applications of Social Network Analysis* (pp. 227–247). Switzerland: Springer.

Rashidi, T. H., Abbasi, A., Maghrebi, M., Hasan, S., & Waller, T. S. (2017). Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies*, *75*(1), 197–211. doi:10.1016/j.trc.2016.12.008

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, *37*(2), 267–307. doi:10.1162/COLI_a_00049

Thackeray, R., Neiger, B. L., Hanson, C. L., & McKenzie, J. F. (2008). Enhancing promotional strategies within social marketing programs: Use of Web 2.0 social media. *Health Promotion Practice*, *9*(4), 338–343. doi:10.1177/1524839908325335

Vuchelen, J. (2004). Consumer sentiment and macroeconomic forecasts. *Journal of Economic Psychology*, *25*(4), 493–506. doi:10.1016/S0167-4870(03)00031-X

Yacob, N., & Mahdzan, N. S. (2014). The predictive ability of consumer sentiment's volatility to the Malaysian stock market's volatility. *Afro-Asian Journal of Finance and Accounting*, 4(4), 460–476. doi:10.1504/AAJFA.2014.067018

Zhou, Q., Xia, R., & Zhang, C. (2016). Online shopping behavior study based on multi-granularity opinion mining: China versus America. *Cognitive Computation*, *8*(4), 587–602. doi:10.1007/s12559-016-9384-x