International Journal of Analysis and Applications

# Determining Consumer Confidence Indexes Using Survey Data and Twitter Data

**Aseel Almansour***

*King Abdulaziz University, Jeddah, Saudi Arabia*

**Corresponding author:* aseel_almansour@yahoo.com

ABSTRACT: The objective of this study was to examine the feasibility of using a new data source—social media (particularly Twitter)—to calculate a consumer confidence index (CCI) and to compare the results with a CCI constructed based on conventional survey data. The CCIs were calculated by collecting data from Saudi consumers through surveys and from social media from January to May 2019. Twitter data were analyzed using two approaches, polarity calculation and an Arabic sentiment database, to reflect the various ways in which the data may be used. The comparison between the survey and social media-derived outcomes may provide insights into the need for—and relevance of—social media data as a better predictor of consumer confidence. Overall, this study highlights the importance of using Twitter data to calculate a CCI.

## 1. Introduction

The world is currently generating new data sources and it is important to embrace these revolutions. In this study, we consider the feasibility of developing a consumer confidence index (CCI) using large amounts of social media data (particularly Twitter) as an alternative for the conventional survey method. CCIs are determined for various use by corporations, governments, and policymakers.

A CCI is calculated using data collected through conventional telephone surveys presented to a sample of the population on a monthly basis. The purpose is to measure the population sentiment on the economic climate, discover whether consumers are optimistic or pessimistic about their financial situation, and determine their spending nature [1] (Shayaa *et al.*, 2018). The consumers are tested on five topics, which translate into the main indicators and sub-indicators

Received Aug. 7, 2021

*2020 Mathematics Subject Classification.* 91B42.

*Key words and phrases.* consumer confidence index; survey data; Twitter data.

https://doi.org/10.28924/2291-8639-24-2026-95



ISSN: 2291-8639

in this study: two related to their current financial situation—consumers are questioned about their current business and employment condition—and three related to their perceptions and expectations regarding the economic conditions in relation to business, employment, and overall family income.

Twitter is a well-known social networking platform where users can freely post their views, opinions, and thoughts [2]. According to Wang *et al.* [3], an accurate estimation of consumer confidence from Twitter data is significant, since it can help establish a broad understanding of consumer purchase behavior. Twitter data are easily available, free, and of low cost [4].

Daas and Puts [5] showed that securing information from social media decreases the response problem and administration fee involved in conventional surveys. The data are analyzed promptly to capture changes in consumer confidence, as high volumes of information are uploaded daily. Social media data have wide coverage since consumers are free to discuss any topic, whereas household surveys only produce responses to predetermined questions. Shayaa *et al.* [1] showed that consumer confidence behavior can be determined through integrated social media big data to complement the traditional methods of consumer confidence evaluation.

According to Arora *et al.* [6], social media provides platforms for users to freely share their reviews and express their opinions on various products or services. The data are defined as "positive" if they contain any positive word and "negative" if they contain any negative word. According to Shayaa *et al.* [1], analytical sentiments included in social media data are essential, allowing consumers to convey their individual characteristics of psychological behavior, reviews of commodities, and feelings toward a social event.

## 2. Methods

This study was conducted over five months, from January 2019 to May 2019, to evaluate the developed CCIs based on survey and Twitter data to capture the trends of the five topics in Saudi Arabia. The main indicators constituted the index of opinions on the economic situation. The sub-indicator sought opinions on job opportunities, home price, house purchase, economic status, economic situation, financial status, house sell, and buying a car.

The analysis of these indicators collected from the survey data was based on Michigan's method of measuring optimistic and pessimistic opinions. The optimistic response rate was measured by the sum of optimistic opinions divided by total opinions; the pessimistic response rate was measured by the total number of pessimistic opinions divided by the total number of opinions. To calculate the weighted index between 0 and 200, the difference between the two ratios were calculated and 100 was added to it, as expressed in the following formula: (positive reviews – negative reviews) + 100.

The data obtained from Twitter were collected based on specific keywords and tweets

analyzed using polarity calculation and the Arab Sentiments database to capture the indicators, following the above methodology for the CCIs. Twenty main keywords were listed, 12 of which concerned economic status, financial condition, jobs, and prices. All keywords were examined separately, and quotation marks were used.

2.1. Methodology of Polarity Calculation for Twitter data

Calculating the polarity of Twitter data involved analyzing opinions expressed from one of these three angles: positive, negative, or neutral. Three advanced models—TextBlob Library, word sequence demystification, and SentiWordNet—were employed. TextBlob Library is a python library that processes text and linguistic data by grammatical tagging, extracting nominal phrases, and analyzing opinions. The word sequence demystification systems treat ambiguity in words by training the machine algorithms for determining semantics. SentiWordNet is a network of linguistic dictionaries based on the relationship between the meanings of words. The method of combining machine learning with the dictionary-based approach known as the lexicon-based method of classification was adopted using the Arab Feelings Database of the National Research Center in Canada.

2.2. The Arab Sentiments Database Analysis Method

This method includes 85,388 positive and negative Arabic words, most of which are in Arabic dialects and generated automatically by Arabic websites and social media. Furthermore, 136,274 words translated from Arabic to English and aligned automatically to evaluate opinions in social communication data, distinguishing between the two languages.

Consumer confidence indicators were analyzed for January, February, March, April, and May of 2019 from the Arab Sentiments database, using the lexicon-based approach. Next, a comprehensive list of positive and negative words incorporated into the dictionary glossary database was prepared. The results were derived by matching the data from the Arab Feelings Database with January tweets, which included the following keywords: economic situation, financial condition, job opportunities, price expectations, house buying or selling, and car buying. Thereafter, optimistic and pessimistic opinions were analyzed using the Michigan approach, according to the administrative regions from 2000; the study extracted the weighted index based on these data for comparison with the published weighted CCI. In this manner, CCIs were calculated using mass media, Twitter, and other online data sources.

## 3. Results

In this section, we present the results of the CCIs calculated from the survey and Twitter data. Table 1 compares the calculated consumer confidence indicators for January 2019.

**Table 1.** Comparison of CCIs from Twitter and survey in January 2019

| January 2019 | CCI from Twitter | CCI from Survey |
|---|---|---|
| Economic situation | 148 | 158 |
| Financial status | 175 | 96 |
| Job opportunities | 176 | 152 |
| Economic status | 80 | 101 |
| Prices | 146 | 118 |
| House purchase | 154 | 88 |
| House sale | 100 | 77 |
| Vehicle purchase | 157 | 87 |

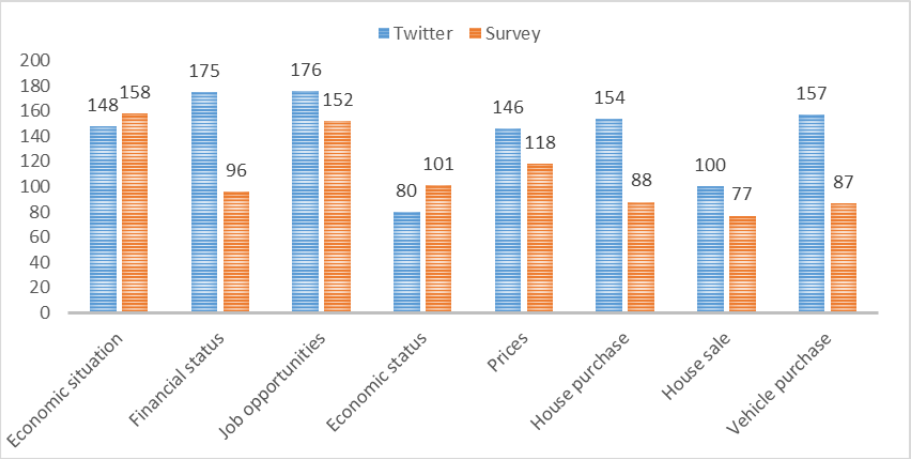


**Figure 1.** Comparison of CCIs from Twitter and survey in January 2019.

As can be observed from Table 1 and Figure 1, the CCI constructed from Twitter data included more responses than the survey-derived CCI, except on two indicators: economic status and economic situation. It was found that for the economic situation indicator, the number of survey confidants was higher than that of Twitter confidants. The financial status indicator showed that the 175 consumers on Twitter had more confidence than the 96 consumers in the survey. A higher number of Twitter confidants were optimistic about buying and selling a house compared to the survey confidants.

Table 2 shows the comparison of the Twitter- and survey-derived CCIs in February 2019. It revealed that for two (25%) indicators (economic situation and job opportunities), survey responses had higher weights than Twitter responses. For all other indicators, Twitter had higher weighting.

**Table 2.** Comparison of CCIs from Twitter and survey in February 2019

| February 2019 | CCI from Twitter | CCI from Survey |
|---|---|---|
| Economic situation | 148 | 156 |
| Financial status | 136 | 133 |
| Job opportunities | 121 | 152 |
| Economic status | 103 | 100 |
| Prices | 136 | 123 |
| House purchase | 130 | 87 |
| House sale | 123 | 80 |
| Vehicle purchase | 118 | 89 |

In February, more Twitter confidants (123) were positive about house sale compared to the survey confidants (n=80).

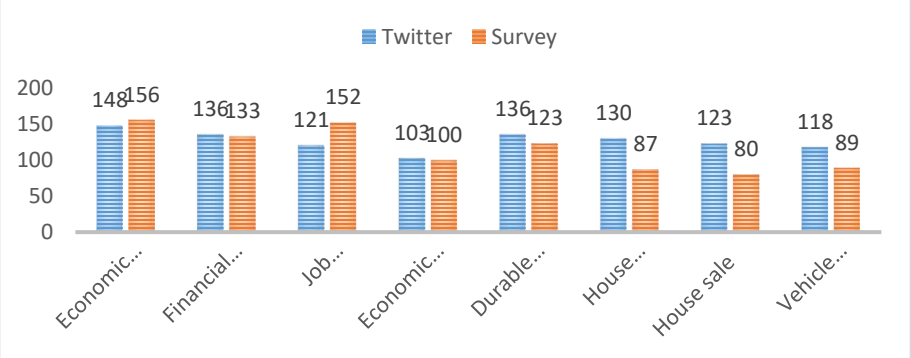


**Figure 2.** Comparison of CCIs from Twitter and survey in February 2019.

As can be observed from Table 2 and Figure 2, compared to the responses in January 2019, those from Twitter confidants declined, indicating consumers' pessimism regarding durable goods prices. The house purchase indicator was higher among Twitter confidants than survey confidants, although both experienced slight deviations from the previous month.

Table 3 presents the comparison of CCIs from Twitter and the survey in March 2019.

**Table 3.** Comparison of CCIs from Twitter and survey in March 2019

| | CCI from Twitter | CCI from Survey |
|---|---|---|
| Economic situation | 134 | 162 |
| Financial status | 133 | 101 |
| Job opportunities | 122 | 148 |
| Economic status | 128 | 105 |
| Prices | 141 | 124 |
| House purchase | 123 | 88 |
| House sale | 126 | 78 |
| Vehicle purchase | 122 | 89 |

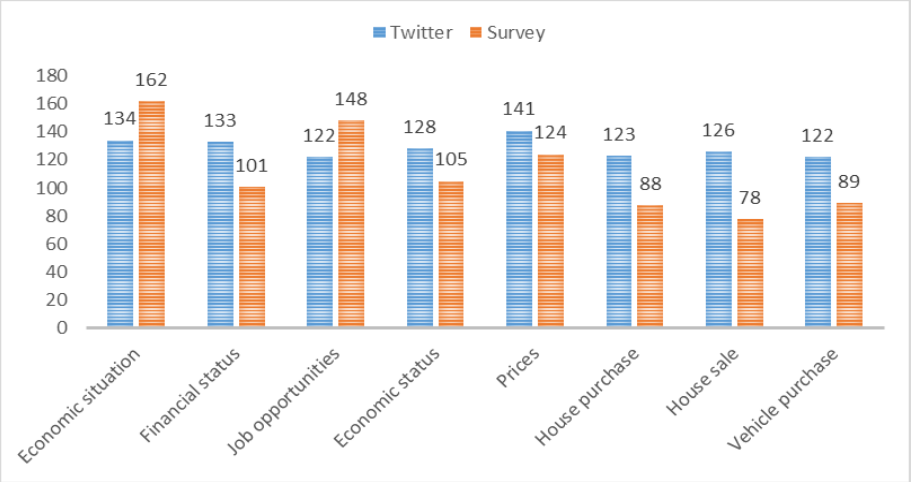


**Figure 3.** Comparison of CCIs from Twitter and survey in March 2019.

Table 3 and Figure 3 reveal that in March 2019, survey confidants were more optimistic than their counterparts on Twitter. For instance, the survey confidants were more optimistic about improving their economic situation compared to Twitter confidants. There was a decline in the number of Twitter confidants who were optimistic about job opportunities in that same month.

During April 2019, both Twitter and survey confidants showed higher optimism regarding the economic status of the country, compared to March 2019 (Figure 4). Both sets of confidants also indicated increasing prices of durable goods, with the Twitter confidants indicating more of a rise.

**Table 4.** Comparison of CCIs from Twitter and survey in April 2019

| | CCI from Twitter | CCI from Survey |
|---|---|---|
| Economic situation | 142 | 162 |
| Financial status | 131 | 103 |
| Job opportunities | 135 | 148 |
| Economic status | 153 | 106 |
| Prices | 135 | 124 |
| House purchase | 134 | 82 |
| House sale | 126 | 85 |
| Vehicle purchase | 131 | 90 |

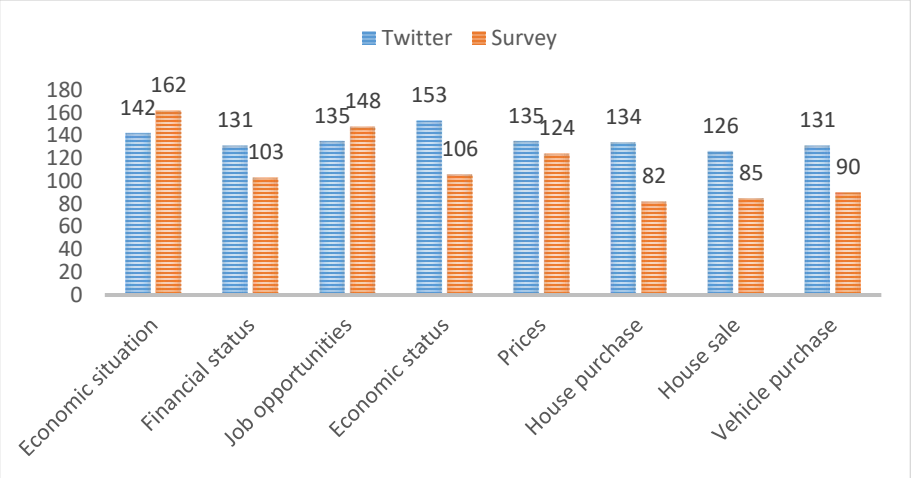


**Figure 4.** Comparison of CCIs from Twitter and survey in April 2019.

As presented in Table 4 and Figure 4, the house purchase indicator showed a positive deviation in both Twitter and survey responses, with a higher indication from the Twitter confidants in that month. In April 2019, there were positive changes in the economic status indicator for both. House sales recorded no change from the previous month among the Twitter confidants; however, a positive deviation was observed among the survey responses.

Table 5 indicates that Twitter confidants had a positive change in their economic situation. They recorded a slight decline in May 2019. The financial status indicator for Twitter confidants recorded a positive change compared to that for survey confidants, whose data indicated a decline that month. The most substantial variation between Twitter and survey data was regarding housing (129: Twitter, 79: survey) and vehicle purchases (132: Twitter, 87: survey). This variation implies that there may have been a general decline in job opportunities or prospects that month.

**Table 5.** Comparison of CCIs from Twitter and survey in May 2019

| | CCI from Twitter | CCI from Survey |
|---|---|---|
| Economic situation | 151 | 156 |
| Financial status | 156 | 101 |
| Job opportunities | 133 | 143 |
| Economic status | 148 | 104 |
| Prices | 112 | 131 |
| House purchase | 129 | 79 |
| House sale | 100 | 81 |
| Vehicle purchase | 132 | 87 |

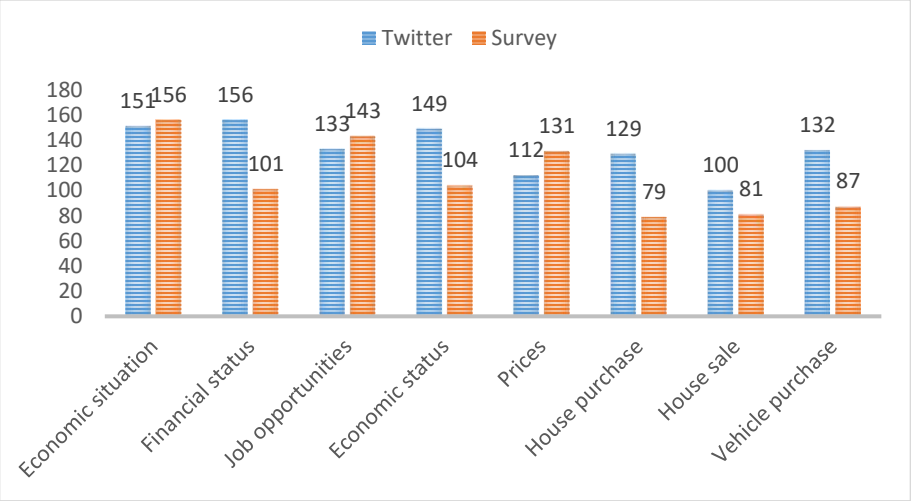


**Figure 5.** Comparison of CCIs from Twitter and survey in May 2019.

In May 2019, the economic status and house sale indicators recorded negative deviations among both survey and Twitter confidants (Figure 5). There was a similar decline in house purchase indicators for both. More Twitter confidants were optimistic about their economic statuses than their survey counterparts. Furthermore, the prices indicator had negative deviations both in the Twitter and survey confidants, compared with the previous month, although the Twitter confidants were more optimistic than the survey confidants.

## 4. Conclusions

CCIs make an important contribution toward a better understanding of economic business cycles and gives an indication of future economic activity. This study investigates the feasibility of calculating a CCI using big data—Twitter—as a novel source. The results presented a better predictor of consumer confidence from Twitter than conventional survey data. In the household surveys, the respondents were asked many questions via telephone, whereas Twitter responses were collected from the social networking platform where users freely post their views, opinions, and thoughts [2]. Due to the high volumes of data continuously being uploaded, social media provides highly relatable information that captures dynamic changes in consumer confidence [7]. The results suggest that using social media platforms, such as Twitter, to measure consumer confidence is more advantageous than conventional surveys because data from social media posts can be reviewed daily to gather responses and examine the actual behavior of households more accurately.

Our results reveal the optimum potential in using social media data to capture Saudi consumers' confidence in various economic situations. Social media can provide a clear image of Saudi's consumer expectations, with regards to their spending on goods and services. This study has potential for further investigation on the use of other data sources to construct indices that

could be alternatives for traditional survey-based methods.

**Conflicts of Interest:** The author(s) declare that there are no conflicts of interest regarding the publication of this paper.

## References

[1] S. Shayaa, S. Ainin, N.I. Jaafar, S.B. Zakaria, S.W. Phoong, et al., Linking Consumer Confidence Index and Social Media Sentiment Analysis, Cogent Bus. Manag. 5 (2018), 1509424. https://doi.org/10.1080/23311975.2018.1509424.

[2] Munna Pandey, Rebecca Williams, Nikita Jindal, Anurag Batra, Sentiment Analysis Using Lexicon Based Approach, IITM J. Manag. IT 10 (2019), 68-76. https://doi.org/10.65301/iitm.2019.10.1.1075.

[3] P. Wang, X. Li, X. Zhan, Y. Zhang, Y. Yan, et al., Building Consumer Confidence Index Based on Social Media Big Data, Hum. Behav. Emerg. Technol. 1 (2019), 261-268. https://doi.org/10.1002/hbe2.156.

[4] J.V.C.E. Igboayaka, Using Social Media Networks for Measuring Consumer Confidence: Problems, Issues and Prospects, Thesis, Université d'Ottawa/University of Ottawa, Canada, (2015). https://doi.org/10.20381/RUOR-4303.

[5] P.J.H. Daas, M.J.H. Puts, Social Media Sentiment and Consumer Confidence, ECB Statistics Paper No. 5, European Central Bank, (2014).

[6] D. Arora, K.F. Li, S.W. Neville, Consumers' Sentiment Analysis of Popular Phone Brands and Operating System Preference Using Twitter Data: A Feasibility Study, in: 2015 IEEE 29th International Conference on Advanced Information Networking and Applications, IEEE, 2015, pp. 680-686. https://doi.org/10.1109/AINA.2015.253.

[7] J. van den Brakel, E. Söhler, P. Daas, B. Buelens, Social Media as a Data Source for Official Statistics; The Dutch Consumer Confidence Index, Surv. Methodol. 43 (2017), 183-210.