



Towards Constructing Egyptian Consumer Confidence Index (ECCI) based on Social Media Sentiments

Mohamed Abo-eleneen*.
PhD, Consultant, Public
Opinion Poll Center,
Information and Decision
Support Center (IDSC), Cairo,
Egypt
mabouhalima@idsc.net.eg

Rojan A. Nafea
Statistician
Public Opinion Poll Center,
Information and Decision
Support Center (IDSC),
Cairo, Egypt
rojanassem@idsc.net.eg

Nada A. Amin
Statistician
Public Opinion Poll Center,
Information and Decision
Support Center (IDSC),
Cairo, Egypt
nadaaly@idsc.net.eg

Dina T. Mansour
Statistician
Public Opinion Poll
Center, Information and
Decision Support Center
(IDSC), Cairo, Egypt
ditarek@idsc.net.eg

Abstract

In an attempt to produce consumer confidence index (CCI) with a lower cost and in shorter time, this paper tries to construct an index based on social media data to capture the rapid changes of opinions, attitudes and sentiments. By comparing the Consumer Confidence Index based on social media index (CCI-SMI) and the traditional CCI it was found that they give the same conclusion about consumer confidence despite their different values. Hence, CCI-SMI can be used as an indicator for consumer confidence immediate changes.

Keywords: Economic, Consumer Confidence, Sentiment Analysis, Social Media

Keywords: Economic; Consumer Confidence; Sentiment Analysis; Social Media.

1-Introduction

Internet technology evolved rapidly in recent years allowing for more interaction between users. Social Media platforms were one of the most important results of this improvement, and with its wide usage around the world, it is considered as main source of data. Now every person can express his or her opinion freely on different social media platforms. The number of its users in 2015 was over 2 billion user accounts around the world across all social media platforms which represent 29% of the world's total population (European Publishers Council, 2015). In Egypt, the percentage of social media users - according to Information and Decision Support Center (IDSC) based on a public opinion poll conducted on a sample of adults 15 years or older in July 2016 - was 29% from total population. The most frequently used social media websites from total population were Facebook, Instagram and Twitter (28%, 5% and 3% respectively).

In the past years, researchers became interested in analyzing sentiments, attitudes and opinions of social media users. Sentiment analysis (SA) is used to classify and categorize different attitudes and opinions that are written, it is usually done by categorizing opinions into three categories positive, neutral, and negative.

SA has many applications such as marketing (it helps marketers to measure consumers' attitudes towards their products), politics (by giving an early alert for potential threats, predicting elections and figuring out people opinions about certain issues), finance (by allowing investors to easily get business news and market information) Duwairi et al. (2014), and tourism (to get travelers' reviews about hotels, trips and countries).

The Organization for Economic Co-operation and Development (OECD) considers CCI as a leading indicator (LI). LI provide qualitative information used to monitor the current economic situation and seek to provide an early warning of turning points in economic activity. It is a tool to measure one of the economic prospects which is Consumers' confidence in economy. Usually CCI is



based on polls conducted by telephone interviews mainly. Nowadays, researchers began to observe social media as an indicator of consumers' confidence, and to analyze Social media users' opinions they began to use SA to categorize those opinions.

SA is usually done by automotive tools. But unfortunately, there are few tools that can deal with Arabic language with high accuracy, as there are many dialects in Arabic language, beside Franco-Arabic (where Arabic words are written in English alphabetic) that is used by many social media users.

In this paper, SA will be used to analyze a sample of tweets retrieved from Twitter by Rapidminer software to calculate Consumer Confidence Index based on social media CCI-SM, and comparing its results with the traditional CCI that is calculated based on telephone interview poll.

2-Related work

Piet J.H. Daas and Marco J.H. Puts (2014) studied in depth the relationship between the traditional CCI and the Dutch social media sentiment to know if there is an ability to use the social media index as a rapid indicator to the changes in the consumer confidence index or not .

They depend on two data sources, the first was consumer confidence data collected by Statistics Netherlands depending on a questionnaire, that consisted of 5 questions each one has three answers positive, neutral and negative. The indicator for each question is calculated by subtracting the percentage of negative answers from the percentage of positive answers. CCI score is the simple average of the results of the 5 questions.

$$CCI = |[\% \text{ of -ve answers} - \% \text{ of +ve answers}]| / 5$$

The second source of data was social media messages; from company that routinely collects public social media messages on the most popular social media platforms. They got around 3 billion messages from 2009 to 2014. The messages were classified as positive, neutral or negative. CCI-SM was calculated as the difference between the percentage of positive and negative messages.

The study proved that there was a correlation between traditional CCI and CCI-SM, so the latter can be published before the traditional CCI.

Another study conducted by Brakel et al. (2016) tried to investigate if social media data and other unstructured data can be used to get official statistics. They used time series modeling to test if data extracted repeatedly from social media and other unstructured data get similar results as repeated surveys or not. CCI-SM was developed from Facebook and Twitter messages written in Dutch, comparing it with series of monthly traditional CCI depending on telephone survey of a sample about 1000 respondents. They conclude that "the traditional *CCI obtain a more stable trend of respondents opinions about the economic climate than CCI-SM*", however the evaluation of both series were the same. Therefore, they suggest that CCI-SM can be used as an indicator for Dutch population attitudes in near real time.

3-Traditional CCI in IDSC

IDSC has started developing a CCI since 2003. This index adopts the methodology of Michigan University in the United States of America. It is based on the results of a monthly landline telephone survey with households. The number of responses ranges from 1000 to 1500 and adjustment weights are applied to represent the actual population in Egypt according to the region and place of residence.

The traditional CCI is the simple average of the three sub-indices which are: The Household Income Level Index (HILI), The Confidence in the Current Economic Policy Index (CCEPI), and The Expectations for the Improvement of the Household and the Country Economic Status Index (EIHCESI) and it depends on eight questions.



Each sub-index is calculated from the results of certain questions, for each question the answers are classified as favorable, neutral and unfavorable. Then, the difference between the percentages of the household heads who gave positive and negative responses is calculated. After that, each sub-index is calculated as the simple average of those differences and the value 100 is added to the final score. Also the general index is calculated as the simple average of the score of three sub-indices.

The value of CCI ranges from 0 to 200 points. If the index value is greater than 100 points that means that the “favorable” responses are more than the “unfavorable” responses. Hence, the consumers are optimistic and tend to purchase more goods and services. This increase in spending inevitably stimulates the whole economy.

On the other hand, if the index value is less than 100 points this means that the “favorable” responses are less than the “unfavorable” responses. Consequently, the consumers are pessimistic and tend to reduce consumption of goods and services. This decreasing in spending reduces the growth of the economy as a whole. Finally, if the index value equals 100 points this means that the “favorable” responses equal the “unfavorable” responses

1. The HILI is calculated from the following questions:

- 1.1 Is the financial condition of you and your family nowadays better or worse or the same as the last year?
- 1.2 Nowadays, Are the prices of durable goods are suitable for your income to buy or not?

2. While the CCEPI's include:

- 2.1 Do you think that economic situation in Egypt nowadays is better or worse or the same as last year?
- 2.2 Do you expect the governmental amendments and decisions will improve your and your family financial condition or it will make it worse or it will be the same as before?
- 2.3 Are the job opportunities available or somehow available or not available at all nowadays?

3. At last, the EIHCESI is composed of:

- 3.1 Do you expect your and your family financial condition will be better or worse, or remain unchanged the next year?
- 3.2 Do you expect that the economic situation in Egypt will be better or worse, or remain unchanged the next year?
- 3.3 Do you expect that the number of job opportunities available will increase, or decrease, or remain unchanged?

4-The suggested CCI-SM

The number of Egyptians who access Social media platforms is increasing notably since 2011. Therefore, a huge amount of data became available through social media. These opinions that are shared or tweeted represent an important data for policy maker, business society and official statistics. Hence, the analysis of such opinions is very useful for many issues; one of these is getting an indicator for consumer confidence in the economy.

In this paper SA is used to analyze the opinions of Egyptians that use Twitter, where users are allowed to write only 140 characters, and it's one of the trendy social media web sites all over the world. Rapidminer software has been used to retrieve the tweets from Twitter and manual categorization for tweets has been made whether they are positive or neutral or negative.



The tweets retrieved according to some keywords that represent the meaning of the eight questions in the telephone survey questionnaire. The following table shows the used keywords.

Table (1): The keywords used to retrieve tweets

Indices and questions	Arabic keyword	Meaning in English
1. HILI		
1.1	الاستهلاك، الحالة المادية	Consumption, financial status
1.2	أسعار السلع	Price of goods
2. CCEPI		
2.1	الاستثمار، إرهاب سيناء، الفساد، المساواة، الإنتاج، الحالة الاقتصادية، التخلص من حكم الإخوان، المدخرات، جودة التعليم، المؤسسات، صندوق النقد، انخفاض الأسعار، الاحتياجات، السياحة، البطالة، عدم الاستقرار، الرقابة والتمويل	Investments, Sinai terrorism, corruption, equity, production, economic situation, Get rid of the rule of the Muslim Brotherhood, savings, the education quality, institutions, Monetary Fund, decreasing price, needs, tourism, Bullying, non-stability, monitor and supply.
2.2	جمارك، العلاوة، المرتبات، القوانين، تعويم الجنيه، الدولار، الضرائب، تغيير الوزراء، سياسة الحكومة، تعليم، الحد الأدنى للأجور، الحد الأقصى للأجور، المستشفيات	Customs, periodical increasing in wages, wages, laws, free float pound, Dollar, taxes, Changing Ministers, government policy, education, minimum wages, maximum wages, hospitals
2.3	الوظائف، فرص العمل البطالة،	Unemployment, jobs, jobs opportunities
3. EIHCESI		
3.1	أطفال، مستقبل أولادنا، المستقبل	Children, future of next generations, the future
3.2	أسعار بكرة	Future prices
3.3	مشاريع جديدة	New projects

5-Results

850 tweets were retrieved for the above keywords during the first two weeks of February 2017, 118 of them were neutral, 219 were positive and 513 were negative. The same methodology of traditional index was used to compute CCI-SM. The following table shows the results of the general index and three sub-indices based on SA of social media tweets compared with the traditional CCI (that was conducted on January 2017 as a last point in the traditional CCI).



Table (2) The Comparison between CCI-SM and traditional CCI.

	CCI-SM	Traditional CCI
CCI	68.6	80.5
HILI	43.3	13.1
CCEPI	51.0	78.1
EIHCESI	111.5	150.3

The values of CCI-SM and traditional CCI (68.6 points and 80.5 points respectively) are lower than 100 points. This means that the “positive” responses are less than the “negative” responses in both methods. Consequently, the consumers are pessimistic and tend to reduce consumption of goods and services.

Moreover, from Table (2) it can be concluded that although the sub-indices about current situation HILI and CCEPI from both methods show that consumers are not confident in the current economic situation, the EIHCESI depending on both methods shows that consumers are more confident about the future economy of Egypt.

The value of CCI-SM is relatively lower than the value of traditional CCI (68.6 points and 80.5 points respectively). The values of CCEPI and EIHCESI sub-indices of CCI-SM are relatively lower than the corresponding values of traditional CCI. On the contrary, the value of HILI sub-index of CCI-SM is relatively higher than the corresponding values of traditional CCI.

The differences in the results between the two methods can be justified by comparing the age and attitudes of respondents. According to a public opinion poll conducted by IDSC on a sample of adults 15 years or older in July 2016, it was found that younger adults (15 – less than 30) tend to use social media platforms more than elder adults (50 years old or older) (52.2% and 5.8% respectively). Hence, when comparing CCI-SM value (68.6 points) and traditional CCI value in the age category 18 – less than 30 (64.4 points); they were relatively equal. In addition, CCI-SM is relatively higher than the traditional CCI due to the fact that the percentage of elder people in TS was more than that of younger.

Limitations & Recommendations:

1. CCI-SM can be enhanced by using other social media platforms.
2. It is better to compare the trend of traditional CCI results and the analyzed tweets over many months in order to get more reliable results, but CCI-SM represents a starting point of a time series that will be tracked.
3. More advanced software for analyzing tweets is needed to guarantee analyzing Arabic messages effectively due to the difficulty of analyzing Arabic language words with its different Dialects and Franco-Arabic words.



Bibliography

- Al-Ayyoub, M., Alsamdi, I., (2014), Lexicon-based sentiment analysis of Arabic tweets, *Int. J. Social Network Mining*, Vol. X, No. Y.
- Al-Harbi, W., Emam, A. (2015), Effect of Saudi dialect preprocessing on Arabic sentiment analysis, *International journal of advanced computer technology*, Vol. 4, No. 6.
- Al-Kabi, M., Gigieh, A., Alsamadi, I., Wahsheh, H., Haidar, M. (2015) An opinion Analysis Tool for Colloquial and Standard Arabic, *Evaluating Association between Research and Social Networks*.
- Al-Subaih, A., Al-khalifa, H., (2014), A System for Sentiment Analysis of Colloquial Arabic Using Human Computation, *Hindawi Publishing Corporation, The Scientific World Journal*, Vol. 2014.
- Brakel, J., Söhler, E., Daas, P., Buelens, B. (2016), *Social Media as a Data Source For Official Statistics; the Dutch Consumer Confidence Index*, Statistics Netherlands.
- Duwairi, R., Marji, R., Sha'ban, N., Rushaidat, S., (2014), Sentiment Analysis in Arabic tweets, *5th International Conference on Information and Communication Systems (ICICS)*.
- Duwariri, R., Qarqaz, I. (2014), Arabic Sentiment Analysis using Supervised Classification, *The 1st international workshop on social networks analysis, Management and security*, Barcelona, Spain.
- Global Social Media Trends (2015)*, The European Publisher Council Global Media Trends Book – Series, Vol. 2.
- Hassan, T., Ali, M., Hedar, A., Doss, M. (2013) Mining social Networks' Arabic slang comments, *IAIDS European conference on data mining 2013 (ECDM'13)*, 22-24 July, Prague, Czech Republic.
- Mao, H., Counts S. (2011), *Computational Economic and Finance Gauges: Polls, search & Twitter*, Paper presented at the Behavioral Economics Working Group, Behavioral Finance Meeting. Palo Alto, CA. Retrieved February 12th, 2017, from <http://www.nber.org/~confer/2011/BEf11/BEf11prg.html>.
- Mao, H., Counts, S., Bolen, J., *Computational Economic and Finance Gauge: Polls, Search, & Twitter*.
- Nabil, M., Aly, M., Atiya, A., (2015), ASTD: Arabic Sentiment Tweets Dataset, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2515–2519, Lisbon, Portugal.
- Piet J.H. Daas and Marco J.H. Puts (2014), *social Media sentiment and consumer confidence*, statistics paper series, European central bank.
- Shukri, S., Yaghi, R., Aljarah, I., Alsawalqah, H. (2015), *Twitter Sentiment Analysis: A case Study in the Automotive Industry*, *IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*.