Received: 22 June 2019 | Revised: 12 October 2019 | Accepted: 28 November 2019
DOI: 10.1002/isaf.1463

RESEARCH ARTICLE

WILEY

Big data tools for Islamic financial analysis

Emna Mnif[1] | **Anis Jarboui**[2] | **M. Kabir Hassan**[3] | **Khaireddine Mouakhar**[4]

[1]Lartige Laboratory, Sfax University, Sfax, Tunisia
[2]ISAAS, Sfax University, Sfax, Tunisia
[3]Department of Economics and Finance, University of New Orleans, New Orleans, LA, USA
[4]EM Normandie, Laboratoire Métis, Le Havre, France

Correspondence
M. Kabir Hassan, Department of Economics and Finance, University of New Orleans. New Orleans, LA 70148, USA.
Email: mhassan@uno.edu

Summary

Behavioural science states that emotions, principles and the manner of thinking can affect the behaviour of individuals and even investors in their decision making on financial markets. In this paper, we have tried to measure the investor sentiment by three means of big data. The first is based on a search query of a list of words related to Islamic context. The second is inferred from the engagement degree on social media. The last measure of sentiment is built, based on the Twitter API classified into positive and negative directions by a machine learning algorithm based on the naive Bayes method. Then, we investigate whether these sensations and emotions have an impact on the market sentiment and the price fluctuations by means of a vector autoregression model and Granger causality analysis. In the final step, we apply the agent-based simulation by means of the sequential Monte Carlo method with the control of our Twitter measure on Islamic index returns. We show, then, that the three social media sentiment measures present a remarkable impact on the contemporaneous and lagged returns of the different Islamic assets studied. We also give an estimation of the parameters of the latent variables relative to the agent model studied.

KEYWORDS
agent-based simulation, big data, Islamic finance, sentiment analysis, sequential Monte Carlo simulation

1 | INTRODUCTION

The number of Muslims in 2010 is estimated at 1.6 billion worldwide, or 23.4% of the world population. Thus, they have more than $800 billion to invest. This capital is currently increasing by about 15% per year. The recent build-up of liquidity in Muslim countries is attracting the attention of stakeholders in the financial market. That is why many analysts agree to study the Islamic financial market to detect the sources of growth and performance and provide returns in the future. *As a part of these studies, this work* aims to analyse the behaviour of the investor by quantifying the emotional component and assessing the factors affecting the performance of Islamic stock markets.

By analysing Islamic finance, it is observed that it is simultaneously ethical and behavioural finance. Indeed, this finance is a compartment of ethical finance because it is characterized by a moral and socially responsible dimension and could meet a need that goes beyond funding (Guéranger, 2009). Islamic finance is based on the respect of Islamic principles. It acts out as a specific filtering material and exclusion sector, which could deflect the behaviour of an investor to one choice or another. This is why we have chosen to address this issue not yet addressed massively by researchers, part of the evaluation framework of the Islamic sentimental dimension and its impact on the Islamic financial market.

Nowadays, Internet technologies are used by companies and individuals more and more. In other words, social media represents an outlet for instant news reflecting accurate information about financial markets.

Social media is then a huge source of information used by governments for elections, companies, and traders to make their decision on the basis of people's attitudes. This approach is called 'sentiment analysis', which is an intelligent process to discover the feelings and emotions of any kind of people. Sentiment analysis is also seen as a field of natural language processing, serving as a tool to classify texts by their polarity into positive or negative sentiment. This research focuses on studying the Islamic financial market and the attitudes towards these kinds of financial assets. For this purpose, we constructed three types of sentiment measure based on big data. The first measurement is built from different social media users and classified by their polarity. The second measure of sentiment is based on search query data. It is employed to analyse the degree of users' attention to the terms related to 'Islamic faith'. In fact, it constitutes a means of detecting emotions and discovering users' attitudes.

The third measure suggests a Bayesian network based on probabilistic classifiers deriving from a machine-learning approach with supervised learning to visualize user attitude and business insights from social media and web users. It also provides an agent-based model founded on the simulation of investor sentiment with sequential Monte Carlo method, as shown in Figure 1. For this purpose, we will first give a brief literature in Section 2. Then, in Section 3, we will try to visualize and measure the sentiment of a special investor category employing 'Google Trends', social media data, and the Twitter API. The agent-based simulation approach will be explained in Section 4, and the results will be discussed in Section 5. Finally, we will conclude by providing some remarks in Section 6.

2 | RELATED WORKS

2.1 | Previous studies on online investor sentiment

Recently, many studies using Google Trends have shown that the search volume for some keywords is related to several phenomena. However, these data are not issued by the whole population.

In the same way, D'Avanzo, Pilato, and Lytras (2017) affirm that the increasing volume of web users does not reflect the entire population. In other words, these queries are related to a scientific community. In our case, we are especially interested in the effect on financial markets, traders, and policy makers' decisions. Moreover, other researchers agree on the predictive power of models using search query data (Stephens-Davidowitz, & Varian, 2015) and sentiment measures based on social media data employed in the study of Ruths and Pfeffer (2014).

The empirical study, visualizing the attention paid by investors through Google Trends and validating the hypothesis of Barber, Odean, and Zhu (2009), was led by Da, Engelberg, and Gao (2011). In their work, they conclude that the negative connotation of the search query volumes implies pessimism in the psychology of the investors and that the optimistic state is deduced from the positive connotation of the search query on the Internet.

Dzielinski (2012) agrees on the use of the search query in building economic uncertainty to predict stock returns and volatility.

As for Zhu, Wang, Qin, and Wu (2012), the Google Trends data are more transparent and reliable than those collected from interviews and questionnaires because of the voluntary and discreet nature of views expressed by Internet users.

Preis, Moat, and Stanley (2013) suggested that the huge data collected from social media visualizing the interaction between Internet users can explain the behaviour of market movements.

Dergiades, Milas, and Panagiotidis (2015) analysed the influence of Twitter, Facebook, and Google Trends on the fluctuation of European financial market variables. By using a multivariate system, they proved a significant effect in the short term mainly to the bond yield differential of the Greek–German state.

D'Avanzo et al. (2017) conducted a study allowing the use of Google Trends and Twitter API data. They built a pipeline interrogating the Twitter sentiment users about new trends. In this pipeline, they constructed a measure of sentiment by means of a machine-learning algorithm based on the naive Bayes method. Their results showed that this approach based on social media sentiment and Google Trends volumes is plausible.

Renault (2017) studied the influence of investor sentiment measured by 'StockTwits' on intraday stock returns He showed that online

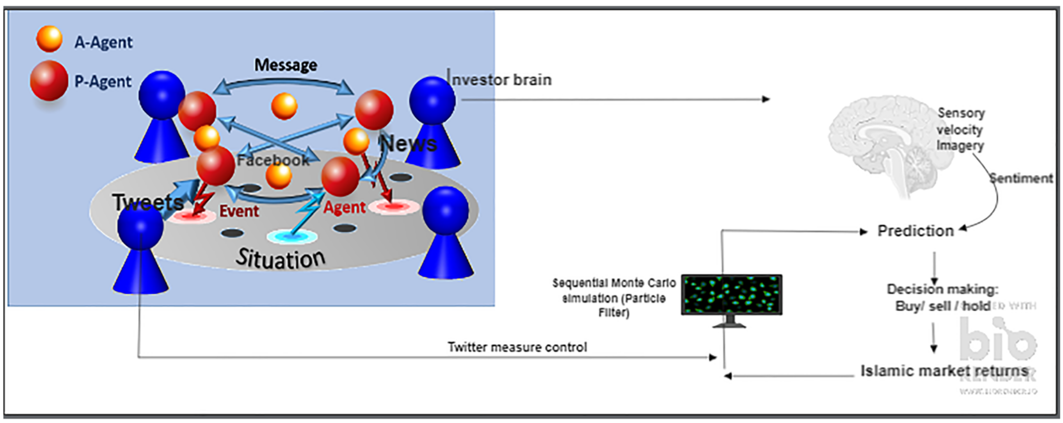


FIGURE 1 Research summary

investor sentiment built by sharing opinions on bullish or bearish indices can help in forecasting their returns.

Tsukioka, Yanagi, and Takada (2018) investigated how investor sentiment, measured by applying the text mining method and the SVM classifiers on the messages data, can affect the performance of the initial public offerings. They found that positive investor sentiment marked by bullish sentiment and that the high attention can positively affect the initial public offering.

Daniel, Neves, and Horta (2016) focused on studying an event's popularity for the 30 companies composing the Dow Jones Average by filtering the tweets published by the financial community influencing financial markets. Their results showed that the constructed measure of Twitter sentiment can detect the events in the associated companies.

Siganos, Vagenas-Nanos, and Verwijmeren (2017) introduced a new measure based on the divergence of sentiment. This Facebook measure is calculated by the distance between people of 20 countries with positive and negative sentiments. Their results showed that the increase in this divergence measure implies more risks and less trading volumes.

Agarwal, Kumar, and Goel (2019) investigated the influence of online information on the market of financial assets through a literature review.

O'Leary (2015) underlined the importance of 'Twitter mining' by synthesizing the extant literature in using Twitter information for prediction, discovery and causation investigation.

Goel and Uzuner (2016) led a qualitative analysis of the text contained in reports for fraud detection by measuring the polarity, intensity, and subjectivity of the sentiment through management discussion and analysis. They found that a fraudulent report contains more positive and negative sentiments than a truthful report.

Fisher, Garnsey, and Hughes (2016) reviewed previous studies in natural language processing application by analysing documents related to the field of accounting, auditing, and finance.

Yu, Duan, and Cao (2013) tested the causality of the sentiment generated by the mentions count of each company through conventional media (newspapers, television, and magazines) and other social media, including Twitter, blogs, and forums, on firm equity value. They found that these different sources of media have an importance on the performance of these companies and can be key in implementing their marketing strategies.

Shen, Urquhart, and Wang (2019) studied the interrelation between the investor attention, measured by the number of tweets, and Bitcoin. They used Granger causality tests to prove these links. Their findings suggest that the tweets count has a significant impact on future Bitcoin market variables.

Unfortunately, empirical studies on the use of big data in Islamic financial markets are scarce. As we know, our work is the first empirical framework focusing on the influence of the psychological component of investors on Islamic assets using big data. Though Muslim people attached to their Islamic religion post positive tweets about the 'Islamic faith', other people who consider Islam as a source of terrorism and injustice write negative tweets. There is a third group of people who do not express any kind of emotions towards 'Islamic faith'. They are, then, exhibiting a neutral sentiment.

2.2 | Agent-based simulation literature review

The estimation of behavioural parameters has been used in many fields, particularly in finance.

Jahangirian et al. (2011) discussed projects focusing on a simulation model's application and management in diverse fields.

In fact, many studies have recently concentrated on simulation-based methods. This is the case, for example, of Guerini and Moneta (2017), who tested the validity of an artificial data simulation model. They compared the vector autoregression of artificial and real data by means of a causality algorithm. In addition, their work has led to the application of a macroeconomic agent-based model.

Alfarano and Milaković (2009) studied the herding behaviour of traders via a probabilistic model based on finance agent simulation.

Franke and Westerhoff (2012), Chen and Lux (2018), and Barde (2016) employed estimated moments in their simulation approach, whereas Kukacka and Barunik (2017) used the maximum likelihood in their simulation method. Diverse estimation methods have been applied by means of the particle filtering approach for state-space models with latent variables. In our case, we are interested in this type of approach, as it is suitable for agent-based models with observable and latent or hidden variables. Lux (2018) estimated two models of agent-based simulation with sequential Monte Carlo estimation via a maximum likelihood online approach estimation and particle Markov chain Monte Carlo. After that, he applied this method to financial data by estimating its parameters via a frequentist maximum likelihood.

Fernández-Villaverde and Rubio-Ramírez (2007) showed how the particle filter approach is useful for likelihood estimation of dynamic macroeconomic models.

In the following, we suggest two hypotheses capable of testing this sentiment influence on Islamic financial market.

Hypothesis 1. *Faith-based sentiment has a positive effect on the lagged Islamic market return, volatility, and the global uncertainty index.*

Hypothesis 2. *Investor sentiment causes Islamic asset returns, volatilities, and economic uncertainty.*

3 | SENTIMENT MEASURES

3.1 | Search query sentiment

In this section we will present the search query tool generated by Google Trends to build a measure of investor sentiment based on terms related to Islamic finance.

In our methodology, the proposed measure is based implicitly on the fact that the data collected from Google Trends presents more transparency, reliability, and accuracy.

This method, which consists of collecting a list of terms related to the emotional context studied, was originally proposed by (Da *et al*. 2011a, Da et al., 2015). Google Trends calculates the count of terms searched by Internet users with Google over time. The trends obtained show the likelihood of a random user to search for a particular search term. Google Trends results indicate a zero value if the search count is low or under the threshold for truncated requests.

The search volume index (SVI) is a normalized measure of the search query count corresponding to the terms submitted to Google. The search volume values vary between 0 and 100 (Narita & Yin, 2018).

This framework exploits Google Trends in order to construct a faith-based sentiment index. For this purpose, we build a list of terms with economic and financial connotations. We follow Tetlock (2007) by using the 'Harvard IV-4' and 'Laswell dictionary of the General Inquirer' in text analysis. Each word is classified by these dictionaries into negative, strong, positive, weak, doctrinal, economic, religious, and so on.

Since we are interested in the faith-based sentiment of the investor, we have selected terms that satisfy both the doctrinal and the economic characters.

However, Islamic finance is a recent field of research whose related words do not exist in such dictionaries or in the classification of 'Google set service', such as: sharia, Islamic principles, ethics, riba, murabaha, Moudharaba, ijtihad, ijma, quardh hassan, Ramadan, musharaka, Quran, Islamic banking, equity funds, short-term investment funds, leasing assets, takaful, Ijara, Ijara waiqtina, istisna'a, bai'salam, bai'muajjal, sukuk, gharar, Mayser, speculation, uncertainty, and so on.

For this goal, we build an initial list of words expressing our sentimental variable formed by combining those terms of Islamic finance and those chosen from the dictionary. We then get the following list:

> Musharaka, profit and loss share, Islamic products, Islamic industry, Islamic trade, Islamic funds, Islamic index, ownership holders, ethical index, Islamic investment, Islamic principles, sharia principles, Islam, ethics, riba, murabaha, Moudharaba, ijthihad, ijmaa, quardh hassan, Ramadan, the Koran, Islamic banking, trust, faith, fate, usury, index performance, hajj pilgrimage, equity funds, short-term investment funds, leasing, takaful, ijara, ijara waiqtina, istisna'a, bai'salam, bai'muajjal, sukuk, donations, alms, alms, Islamic finance, and so on.

After that, we introduce each term and compare their frequencies. A final list of five terms is then retrieved that includes these terms 'Sukuk', 'Islamic index', 'Islamic faith', 'Islamic banking', and 'Islamic finance' in the context of 'finance' in the entire world since 2004 in monthly frequencies.

Though sentiment measures based on surveys show a clear theoretical link to investor confidence, they, unfortunately, take time in the polls and create an offset (Qiu & Welch, 2004). Besides, answers of respondents are frequently biased and insincere (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). That is why we follow Da et al. (2011) to build a measure based on Internet research sense.

The faith sentiment index is built by using the SVI of the final list of keywords generating the most frequent terms related to 'Islamic faith' through Google Trends. Then, we convert these five search terms (SVI—'Sukuk', 'Islamic index', 'Islamic faith', 'Islamic banking', and 'Islamic finance') into reasonable indications by generating the log of each SVI. We use the log of the SVI to summon the time-series changes and to be helpful for interpretation of regression.

The next step is to compute our sentimental index; we then calculate the average of the sum of all SVI search terms related to religious feelings (Srelig):

$$\mathrm{relig}_t = \frac{1}{\theta(\mathrm{Srelig}_t)} \sum_{j \in \mathrm{Srelig}} \log\left(\mathrm{SVI}_{j,t}\right) \tag{1}$$

In this equation, $\theta()$ is an accounting function that recognizes the valid values of the search terms linked to religious feeling found on the day t. This function eliminates the SVI values that are equal to zero when the queries are not frequent for such a daily term.

3.2 | Social media sentiment

As social media differs from traditional media, it becomes another way of communication used by companies, investors, and governments to achieve their goals. For this reason, the practitioners try to apply different methods capable of measuring the engagement of customers, investors, and citizens.

A social media network is a way to express opinions formed by interactions such as 'likes', 'posts', and 'tweets' issued by their users. These are generally used to measure customer engagement (Oviedo-García, Muñoz-Expósito, Castellanos-Verdugo, & Sancho-Mejías, 2014).

We present the ratio of engagement as defined by Muñoz-Expósito, Oviedo-García, and Castellanos-Verdugo (2017) as

$$\text{Ratio of engagement} = \frac{\dfrac{\dfrac{\text{interactions}}{\text{number of tweets}}}{\text{average impressions}}}{\text{average reactions}} \tag{2}$$

As for Oviedo-García et al. (2014), the ratio of engagement on Facebook can be formulated as follows:

$$\text{Ratio of engagement} = \frac{\dfrac{\dfrac{\text{likes + comments + shares + other clicks}}{\text{number of posts}}}{\text{average impressions}}}{\text{average reach}} \tag{3}$$

Brand24 offers essential knowledge about marketing performance measured by social media metrics such as engagement rate, number and type of interactions, social media reach, and social media shares. It also uses an algorithm to measure and classify sentiment into positive and negative. Although, the algorithm cannot always determine context because it has not been able to work around sarcasm, most sentiment analysis results are accurate.

In this study, we try to visualize the attitude of social media users towards Islamic finance. For this goal, we extract our data from Brand24 by introducing the following terms: Islamic faith, Sukuk, Islamic finance, Islamic index, and Islamic banking. Our choice of these terms is related to Google Trends queries. We first evaluate the interest intensity for each social media towards these words summarized in Table 1.[1]

Table 1 indicates that the frequency rate of the keyword 'Islamic faith' in Facebook interaction has the highest level. Facebook is followed respectively by Twitter, videos, blogs, web, Instagram, forum' and finally by the news.

3.3 | Twitter sentiment

This framework is based on two principal goals: of visualizing the attitudes expressed in tweets and of building a sentiment measure.

The sentiment detection module employs a Bayesian algorithm as a sentiment analysis tool. The data set contains 31,530 terms that have been trained on Wiebe's subjectivity lexicon and classified into 'positive' or 'negative' sentiment (Riloff & Wiebe, 2003). This module specifies the feelings expressed by the associated tweet by means of a naive Bayes learning algorithm and classifies the emotions into disgust, anger, surprise, fear, sadness, and joy trained on Strapparava and Valitutti's (2004) **`emotions`** lexicon.

In this section, we analyse the word frequencies for English tweets about 'Islamic faith'. Figure 2 shows the emotion visualization, the polarity, and the word cloud of the term 'Islamic faith' between 15 December 2017 and 19 February 2018.

These figures address the thinking style of Twitter users who have a global positive view towards 'Islamic faith'. The dominant emotion towards this term is 'joy'.

In the next step, we try to define a Twitter sentiment indicator of 'Islamic faith' using the number of tweets. After retrieving the positive words and the negative ones, we build our twitter measure as Twitter Sentiment (TWS)[2]:

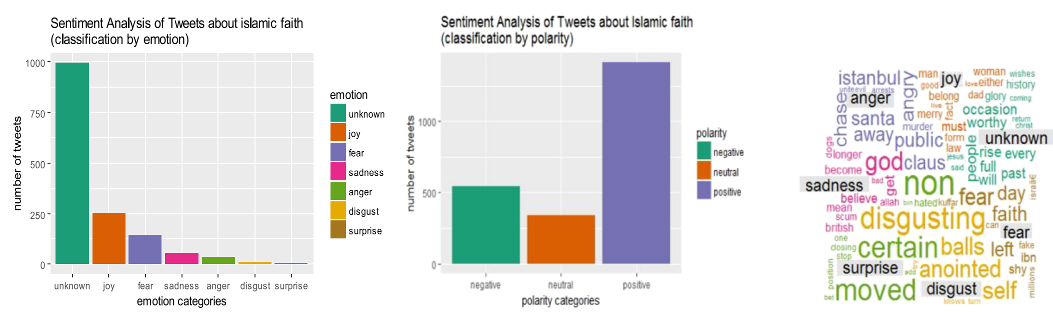
$$\text{TWS} = \frac{\text{positive} - \text{negative}}{\text{positive} + \text{negative}} \tag{4}$$

4 | AGENT-BASED SIMULATION

In recent studies, the simulation methods are based mostly on simulated methods of moments estimation. Markov chain Monte Carlo methods are used to estimate agent-based models for models with latent variables or state-space models.

TABLE 1 Social media overview

Source	Sukuk	Islamic faith	Islamic finance	Islamic index	Islamic banking
All social media	2682	2168	2378	327	2357
Facebook	24	966	252	13	361
Twitter	702	547	735	198	658
Instagram	762	59	283	0	117
Blogs	30	62	54	8	104
Forum	1	46	24	0	32
News	72	36	79	21	69
Video	503	249	690	5	696
Web	367	203	261	82	320



(a) (b) (c)

FIGURE 2 Islamic faith visualization by (a) emotion, (b) polarity classification, and (c) the relative word cloud

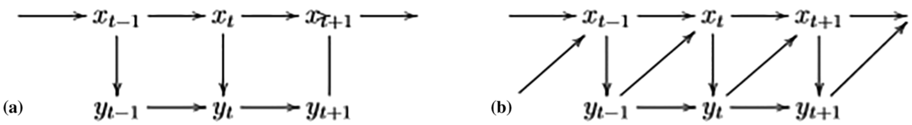


FIGURE 3 Dynamic model with latent states.
Source: Lux (2018) [Colour figure can be viewed at wileyonlinelibrary.com]

4.1 | The sequential Monte Carlo simulation approach

The stochastic evolution over time of the unobservable variables associated with the state vector x_t, and the observable points corresponding to the vector **y** defines the state-space models.

Each variable can be written with the following equation.

$$x_t = f(x_{t-1}, \varepsilon_t) \text{ and } y_t = y(x_t, \eta_t) \tag{5}$$

where ε_t and η_t represent the correlated noise component. The functions *g*() and *f*() are stochastic processes. The state-space model can be diagrammed by the conditional densities of x_t and y_t. In other words, we might use some algorithms to simulate these variables.

The state-space models can be associated with the sequence of events illustrated in Figure 3a. Figure 3b explains the relations between the unobservable variables x_t and the observable variables y_t.

Agent-based models can often follow this cliché because of the complex and unobservable characteristics of agents, especially on financial markets.

4.2 | Agent-based model presentation

In this section we will discuss the prototype model employed in this study and the particle filter algorithm used for the estimation of their parameters. Our model studied is derived from the proposed agent model with sentiment dynamics built by Alfarano, Lux, and Wagner (2008) and estimated by Lux (2018). The price fluctuations are formed by sentiment and fundamental news. We define the evolving log price as

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \beta\left[T_\mathrm{f}\left(p_{\mathrm{f},t} - p_t\right) + NT_\mathrm{c}x_t\right] \tag{6}$$

where $p_{\mathrm{f},t}$ is the current log fundamental value, p_t is the log market price at time *t*, x_t is the sentiment variable, T_f and T_c are parameters that symbolize the total trading volume associated with the two types of traders (fundamental and chartists respectively), *N* is the per-head measure multiplied by the non-fundamental traders' number, and β is the price adjustment speed.

If we adjust prices to clear the market, β retreats as a free parameter and the equilibrium current price becomes

$$p_t = p_{\mathrm{f},t} + N\frac{T_\mathrm{c}}{T_\mathrm{f}}x_t \tag{7}$$

Consequently, returns r_t can be divided into sentiment innovations and fundamental news:

$$r_t = p_{t+1} - p_t = p_{\mathrm{f},t+1} - p_{\mathrm{f},t} + N\frac{T_\mathrm{c}}{T_\mathrm{f}}(x_{t+1} - x_t) \tag{8}$$

We suppose that the log fundamental news follows standard Brownian motion:

$$p_{\mathrm{f},t+1} - p_{\mathrm{f},t} = \sigma_\mathrm{f}\varepsilon_{\mathrm{f},t} \qquad \varepsilon_{\mathrm{f},t} = N(0,1) \tag{9}$$

In this study, we retrieve filtered state probabilities of sentiment as it is considered as a latent variable.

The model of Alfarano et al. (2008) explains the sentiment evolution over time. The module used in this work introduces the big data measure of sentiment as a known control variable in the transition state of the unobservable sentiment variable. For this reason, we employ a bootstrap particle filter, which is a part of the sequential simulation methods of Monte Carlo. This method attempts to approximate the probability distributions of the unknown variable by using simulations. Standard particle algorithms are based on the succession of prediction, correction, and resampling steps that maintain a weighted particle system approaching the posterior law. During the prediction stage, the particles are propagated according to the equation dynamics. Their weights are adjusted according to their adequacy with current observation through the evaluation of the likelihood function. Resampling consists of favouring particles whose weights are significant to conserve an approximation of the non-degenerate posterior law. Another problem raised by particle filtering is the automatic and sequential detection of the divergence which makes it possible to check the integrity of the reset solution provided by the filter and possibly reinitialize it. Following notation in the work of Lux (2018), we denote by $\{r_t\}$, $t = 1,\ldots,T$, the observed variable, by $\{x_t\}$, $t = 1, \ldots,T$, the hidden variable, and by θ the vector of possibly unknown parameters; the entities of interest might be the predictive density $p(x_t | R_{t-1}, \theta)$, the filtering density $p(x_t | R_t, \theta)$, or the smoothing density $p(x_t | R_t, \theta)$, where R_t is the set of all observations from r_1 to r_t. These densities are approximated by the bootstrap particle filter, which is also called sequential Monte Carlo or the multinomial resampling method.

We follow the principles laid out by Kitagawa (1996) in the algorithm implementation. We explain the details of the particle filter on the basis of the Alfarano et al. (2008) model. Ghonghadze and Lux (2016) and Jang (2015) had difficulties in estimating $T_\mathrm{c}/T_\mathrm{f}$ and *N*

[2]Suppala and Rao (2019) detailed the methodology used to measure Twitter sentiment.
[2]Suppala and Rao (2019) detailed the methodology used to measure Twitter sentiment.

because of their discrete nature and the near collinearity with other variables. In this work, we set the same values fixed as Ghonghadze and Lux (2016) and Jang (2015) at $T_c/T_f = 1$ and $N = 100$.

5 | EMPIRICAL INVESTIGATION

5.1 | Data and methodology

For the purpose of studying the return–sentiment and the risk–sentiment relationship, we consider different types of market assets composed of Islamic unit trusts, Islamic bonds or Sukuk and Islamic indices. We calculate the corresponding return and volatility proxies for each Islamic unit trust, Sukuk-market and stock-market index. For instance, Islamic faith sentiment is measured by three different methods and extracted from different data. We have, then, different sources of data as described in the following.

5.1.1 | Financial data

Our financial data are composed of different types of Islamic assets extracted from the Thomson Reuters and Datastream database. These assets are composed of three families. In the first group trained by an Islamic unit trust, we choose the Hang Seng Islamic China Index Fund and the HSBC Islamic Global Equity Index. The second family of Islamic indices is composed of the Nigerian index (NSE Lotus Islamic), the Qatari index (QE Al Rayan Islamic), the British index (Russell-IdealRatings Islamic Dev), the American index (DJIM), the Indonesia index (Jakarta SE Islamic) and the Turkish index (TR Crescent Aus Islamic). The last group is associated with the Sukuk family containing Noor Sukuk Co. Ltd (2015), QIB Sukuk Ltd (2015), AHB Sukuk Co. Ltd (2017), Petronas Glb. Sukuk (2015), Hong Kong Suk (2014), Saudi Elty. Glb. Suk (2012), Equate Sukuk Spc (2017), Avrist Ada Sukuk Berkah Syariah, Jany Sukuk Company (2014). We also use the volatility data extracted from the Global Economic Uncertainty Index.[3]

5.1.2 | Social media data

We collect our social media data from three different databases in order to measure our three corresponding sentimental variables.

The first online data is extracted in daily frequencies during the period between 21 October 2017 and 19 March 2018 from Google Trends, which is a free-access database. The second type of data is retrieved from Brand24, which is a commercial database, during the period between 21 October 2017 and 19 March 2018 in daily frequencies. Finally, the last kind of data is grouped from Twitter API between 15 December 2017 and 19 March 2018 with instantaneous frequencies reduced in daily frequencies by means of moving average (Corea, 2016).

[3]http://www.policyuncertainty.com/global_monthly.html.

5.2 | Methodology

In this section we discuss the methodology used to show the link and effect of our sentimental variables on the daily returns and volatilities of our assets being studied. The measures obtained of Twitter sentiment (TWM) are averaged on each day. In other words, the daily Twitter sentiment (TWD) follows the following formula:

$$\mathrm{TWD}_i = \frac{1}{n}\sum_{t=1}^{n}\mathrm{TWM}_i \tag{10}$$

In this study, we use models with latent variables because the financial agent models are generally characterized by latent variables, presenting non-Gaussian fluctuations and nonlinear dynamics since they reflect the complex reality. For this reason, we propose to use the sequential Monte Carlo rather than Kalman filter, hidden Markov models, Markov chain Monte Carlo, and extended Kaman filter. In this approach, we approximate the probabilities that are involved in the likelihood function of asset returns of Islamic financial data (the observable series) and the agents' sentiment (unobservable variables).

We use the likelihood function to approximate particle filter simulation denoted by:

$$L(\theta) = P(r_1,\ldots,r_T|\theta) = P(r_1|\theta)\prod\nolimits_{t=2}^{T}P(r_t|r_{t-1},\theta) \tag{11}$$

By observing the link between the observable variable (returns) and the state variable (sentiment) in equation (11), the conditional probabilities can be formulated as

$$P(r_t|r_{t-1},\theta) = \int P(r_t\,|x_t)P(x_t|r_{t-1})\,\mathrm{d}x_t \tag{12}$$

$P\,(r_t|x_t)$ obviously is a Gaussian density. x_t and its distribution in equation (12) will be approximated by a set of 'particles' B. Accordingly, denoting the B particles by $x_t^{(j)}$, we approximate equation (12) by

$$P(r_t|r_{t-1},\theta) \approx \frac{1}{B}\sum_{j=1}^{B}p\left(r_t/x_t^{(j)}\right) \tag{13}$$

The steps of the filtering algorithm occur as follows:

1.A random collection of B particles is created at time $t = 1$ and the densities $P\left(r_1|x_1^{(j)}\right)$ are computed.

2.Particles are resampled using weights

$$\frac{P\left(r_1|x_1^{(j)}\right)}{\sum_{t-1}^{B}\left(r_1|x_1^{(j)}\right)} \tag{14}$$

3.The new set of particles obtained is iterated to form a new sample using the stochastic process defined by equation (11). This process can be simulated exactly as a sequence of Poisson events for each particle, and the resulting value of this particle after a unit time step is recorded, which we denote as $x_2^{(j)}$.

4.Steps 1 to 3 are repeated for $t = 2,3,\ldots,T$.

5.3 | Empirical results

5.3.1 | Relation between market variables and investor sentiment

In this section, we try to give the results of testing our hypotheses. For this reason, we check the stationary features of our time series. We use the regression with Newey–West standard errors to avoid any autocorrelation and heteroscedasticity of the errors. Table 2 provides the results of this regression. In this table, we clearly show that most of the coefficients of Twitter, Google Trends measure, and engagement sentiment are positively and statistically significant with the return on Islamic assets. Furthermore, the corresponding positive coefficients for each index are relevant. These results lead us to confirm Hypothesis 1. The second part of Table 2 shows negative significant statistics between the sentimental proxies and volatility. In other words, the increase in the measured sentiment induces a risk decrease. This result confirms our Hypothesis 2.

For our second hypothesis, Table 2I[4] expresses the results of Granger causality test with different lags for Sukuk market. For each Sukuk, unit trust, and Islamic index, we first determine the optimal lag length. In other words, we try to focus on the lagged effect of our sentimental proxies on the return and the volatility of the different Islamic assets on the one hand and the global uncertainty on the other hand. In this Table 3, most of the p values are under 5%, which shows that our sentimental proxies Granger causes the different Islamic assets studied consisting of Sukuk, Islamic unit trusts, and Islamic index returns and market volatility.

Table 3 illustrates the results of the Granger causality test with different lags for unit trusts and the Islamic index market. Table 3 clearly shows that most of the unit trusts and Islamic index are Granger caused by Twitter sentiment, search query sentiment, and social media sentiment.

5.3.2 | Agent-based simulation results

This section analyses the interactions between the investor and the environment by investigating the connection between the brain's ability to predict and the higher cognition generated by the information. In particular, we explore the connection between prediction and mental imagery (Moulton & Kosslyn, 2009).

TABLE 2 Newey–west standard errors regression results for Sukuk

	Search query sentiment	Engagement sentiment	Twitter sentiment
Sukuk			
QIB	−0.0000596 (−0.0000136)	−0.0002299 (-4.94×10^{-8})	0.0003314 (-3.90×10^{-7})
AHB	0.0000158*** (-7.71×10^{-7})	0.0001016*** (-2.27×10^{-8})	-3.16×10^{-6} (-4.13×10^{-8})
Noor	0.0019624** (-9.09×10^{-6})	−0.0001409 (-5.07×10^{-7})	0.0003188** (−1.07 × 10−7)
Petronas	0.0063092** (-5.02×10^{-6})	−0.0000707 (1.04×10^{-7})	−0.0001758 (-4.55×10^{-7})
Hong	−0.0002341 (-7.63×10^{-6})	−0.0001083 (-6.42×10^{-7})	0.0002329** (-2.33×10^{-7})
Saudi	0.0003904** (-2.19×10^{-6})	−0.0002671 (1.40×10^{-6})	0.0005361** (-4.90×10^{-6})
Jany	0.0002744*** (-5.85×10^{-6})	−0.0000901 (3.15×10^{-7})	0.0002512** (-9.91×10^{-7})
Avrest	−0.006348 (9.89×10^{-6})	−0.0000338 (1.04×10^{-6})	0.0006246** (−0.0000152)
Equate	0.0001196** (−0.0000481*)	−0.0002739 (2.17×10^{-6})	0.0011642** (-9.19×10^{-7})
Hang Seng	−0.0478396 (−0.0002438)	−0.003229 (0.0000676)	0.0047123 (−0.0001235)
HSBC	−0.0053462 (−0.0007925**)	0.0001993 (−0.0000767)	0.0019662 (−0.0000978)
Islamic indices			
NSE	−0.0029972 (−0.0009667***)	0.0027417 (0.0000464)	0.003086*** (2.20e−06**)
QE	0.0207615 (−0.0013762)	−0.0033242 (−0.0000238)	−0.0016588 (0.0000315*)
Russell	−0.0152144 (−0.0008427**)	0.0030301 (−0.0000354)	0.0018004** (−0.000087**)
DJIM	−0.0208827 (−0.0008254*)	0.0025703 (−0.0000253)	0.0016885*** (−0.000078*)
Jakarta	0.0386585 (−0.0011331**)	0.0015849 (−0.0000202)	0.0043091* (−0.0000439*)
TR	−0.0768166** (−0.0006214*)	0.0008972 (0.0000298)	0.0016951*** (−0.000068**)
Uncertainty index	−708.6591*	−4.706255	22.43864***

*, **, *** indicate statistical significance at 10%, 5%, and 1% level respectively. The values in parentheses are relative to the corresponding volatility estimation.

TABLE 3 Granger causality results

	Twitter sentiment → index	Engagement sentiment → index	Search query sentiment → index
Sukuk			
QIB	0.011 (0.945)	0.123 (0.005)	0.742 (0.132)
AHB	0.090 (0.076)	0.046 (0.000)	0.312 (0.723)
Noor	0.000 (0.859)	0.002 (0.005)	0.736 (0.162)
Petronas	0.000 (0.000)	0.150 (0.013)	0.666 (0.150)
Hong	0.038 (0.087)	0.008 (0.194)	0.001 (0.018)
Saudi	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Jany	0.000 (0.011)	0.556 (0.000)	0.000 (0.004)
Avrist	0.000 (0.000)	0.001 (0.004)	0.014 (0.013)
Equate	0.000 (0.000)	0.010 0.001	0.000 (0.007)
Unit trust			
Hang Seng	0.037 (0.131)	0.038 (0.000)	0.499 (0.000)
HSBC	0.009 (0.034)	0.150 (0.13)	0.382 (0.001)
Islamic indices			
NSE	0.013 (0.001)	0.000 (0.000)	0.001 (0.001)
QE	0.000 (0.000)	0.001 (0.000)	0.000 (0.000)
Russell	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
DJIM	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Jakarta	0.001 (0.000)	0.008 (0.005)	0.000 (0.000)
TR	0.008 (0.007)	0.010 (0.001)	0.000 (0.000)
Uncertainty index	0.006	0.001	0.001

The values in parentheses are relative to the *p*-value of the corresponding returns and volatility.

In this empirical analysis, we use the Twitter sentiment measure and the Jakarta Islamic index, as they prove the most important interaction and the most significant results (Tables 2 and 3).

As we mentioned in the previous section, investor or trader sentiments are represented by the hidden variables x_t and the returns are represented by the observable variables denoted by y_t.

By fixing $T_c/T_f = 1$ and $N = 100$, equation (9) becomes

$$r_t = p_{t+1} - p_t = p_{f,t+1} - p_{f,t} + 100(x_{t+1} - x_t)$$

Then, equation (10) becomes

$$p_{f,t+1} - p_{f,t} = \sigma_f \varepsilon_{f,t} \qquad \varepsilon_{f,t} = N(0,1)$$

Our goal is to predict returns y_t from the sentiment, which is a latent variable, given noisy measurements. A generative model of its measurements is required and therefore is given by

$$y_t \sim N\left(x_t, \sigma_y^2\right)$$

[4]The symbol 'sentiment measure → return (or volatility)' indicates the null hypothesis of sentiment measure cannot Granger cause returns (or volatility) of indices.

For the dynamics of the latent state x_t, we specify the model by following the study of Stahl and Hauth (2011), which is given by

$$x_t = h\left(x_{t-1} + T_{t-1}, \sigma_x^2\right)$$

T_{t-1} represents the Twitter measures of sentiment and a 'known' control input at the same time. σ_x^2 is the sentiment standard deviation. These terms encode the agent's decision in the process model. The control input is assumed to have been seen by the trader and represent knowledge about the most fundamental news. This situation is thus equivalent to a 'known' covariate that we can use to predict returns.

In the first step (t_0), we choose parameters for the filtering algorithm and the process model (transition model). After that, we make an initial guess about x_t. This means that we draw a sample from the initial distribution; in our model, we assume that the agent starts in an unbiased state with a high degree of certainty. We predict the state at the next time increment by applying the process model to each of the N particles.

In Markov chain Monte Carlo terminology, the predicted states represent samples from a proposal distribution.

At t_1, there are return measurements available, so the agent can use these in order to update the internal estimate of their state. To do so, the agent evaluates the likelihood of each particle. In other words, the agent evaluates the probability of observing the measurement at time t given the predicted sentiment. The probabilities are then used as importance weights for each particle. Intuitively, this means that the agent selects those particles that have predicted the data well and discards those particles that have predicted the data poorly. This is achieved by weighted sampling with replacement from the set of particles (multinomial resampling). The weights are first normalized, so that they represent a probability distribution.

By applying this simple procedure recursively, the agent can track the latent sentiment, given the sequence of noisy returns measurements. At any time step, the agent thus performs the following sequence of computations:

1.Prediction (propagation) step: predict the state one step ahead.

2.Update step: obtain importance weights from return measurements and select particles for 'survival'.

3.Compute summary statistics, which are mean and variance of estimated state.

Figure 4 shows a visualization of the computations performed at each time step: the propagation or the prediction step, the update state, and the summary statistic measurements. After the prediction step, the set of particles represents the new belief before 'seeing' the new evidence. All particles thus have equal weights, given N. Subsequently, the probability of observing the evidence is evaluated for each particle, giving new weight. Each particle's size reflects its importance in predicting the new measurement. The final step consists of sampling 100 new particles with replacement from the set of particles, with a probability of being chosen by their importance weights (this is the multinomial resampling part). Particles that did not predict the data well may, therefore, be eliminated.

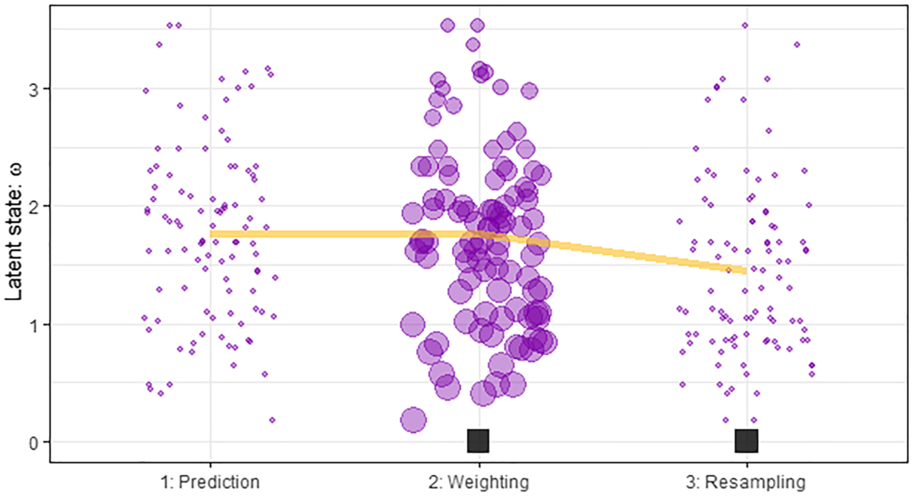


FIGURE 4 Visualization of the computations

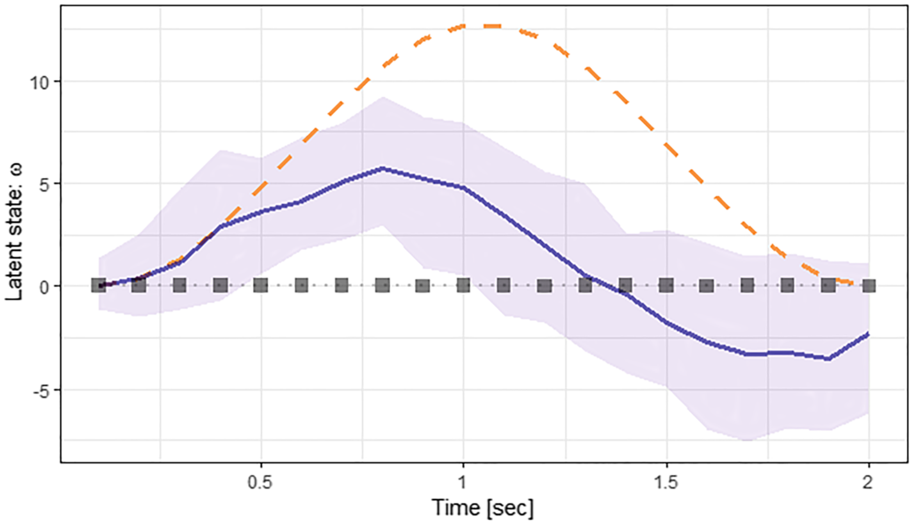


FIGURE 5 simulation of investor sentiment (latent variable)

In Figure 4, the purple spheres show the estimated mean of the distribution being 'shifted' during a single time increment, and the black rectangle shows the observations.

At the first step, before seeing the new data, the observation is at the previous value. At the updating step (step 2), a new measurement is available. The closer to the measurement, the larger the diameter of the particle.

Figure 5 shows the agent (investor or trader) engaged in real-time sensory inference during a 2 s motion event. The dotted line shows the true measurement of sentiment x_t, the black rectangles are the return's measurements, the solid line shows the agent's estimated sentiment, and the bands surrounding the solid line reflect the 95% credible interval.

6 | CONCLUSION

In this paper, the sentiment index based on social media is first constructed according to the comments and interaction data extracted from different social media. Then the top methods are proposed to investigate whether these sensations and emotion have an impact on the market sentiment and the price fluctuations. For this purpose, we analysed the investor behaviour by means of big data extracted from diverse social media, the search engine Google Trends, and the Twitter API. Therefore, we tried to measure the investor sentiment by three means of big data.

The first is based on a search query of a list of words related to Islamic context from 21 October 2017 to 19 March 2018 in daily frequency. The second is deduced from the engagement degree on social media collected by a social networking site called Brand24 from 21 October 2017 to 19 March 2018 in a daily database. The last measure of sentiment is built using a machine-learning algorithm and the Twitter API from 15 December 2017 to 19 March 2018. This sentiment is then visualized by representing the emotions and the polarity of some specific words introduced by Twitter users. In the next step, we estimate the contemporaneous effect by Newey–West regression in order to avoid autocorrelation and homoscedasticity of errors. After that, we focused on the lagged effect of these sentimental proxies by means of a vector autoregression model. Finally, Granger causality analysis was used to investigate the hypothesis that Islamic faith sentiment causes the volatility and the returns on Islamic assets.

Our results show that, on the first level, Twitter users of the term 'Islamic faith' express a more positive sentiment. Second, the three social media sentiment measures have a significant effect on the contemporaneous and lagged returns of the different Islamic assets studied. Using Granger causality, all these sentimental variables also affect the Islamic assets returns and volatilities. We also find that our sentimental variables have a positive effect on the economic uncertainty measured by global economic uncertainty.

In this paper, we have used sequential Monte Carlo based on a bootstrap particle filter. We also used the maximum likelihood estimation for this approximation in the simulation of an agent-based model.

We have also included our sentiment measure based on the Twitter API as a latent variable. This empirical work is, to our knowledge, the first study using big data in the simulation of Islamic financial data with the sequential Monte Carlo method.

As it is a wide field of study, we let future empirical works focus on the other methods of the estimation of the parameters in agent-based simulation.

ORCID

M. Kabir Hassan https://orcid.org/0000-0001-6274-3545

REFERENCES

Agarwal, S., Kumar, S., & Goel, U. (2019). Stock market response to information diffusion through internet sources: A literature review. *International Journal of Information Management*, *45*, 118–131. https://doi.org/10.1016/j.ijinfomgt.2018.11.002

Alfarano, S., Lux, T., & Wagner, F. (2008). Time variation of higher moments in a financial market with heterogeneous agents: An analytical approach. *Journal of Economic Dynamics and Control*, *32*(1), 101–136. https://doi.org/10.1016/j.jedc.2006.12.014

Alfarano S, Milaković M. 2009. Network structure and *N*-dependence in agent-based herding models. *Journal of Economic Dynamics and Control 33*(1): 78–92. https://doi.org/10.1016/j.jedc.2008.05.003

Barber, B. M., Odean, T., & Zhu, N. (2009). Do retail trades move markets? *Review of Financial Studies*, *21*(1), 151–186. https://doi.org/10.1093/rfs/hhn035

Barde, S. (2016). Direct comparison of agent-based models of herding in financial markets. *Journal of Economic Dynamics and Control*, *73*, 329–353. https://doi.org/10.1016/j.jedc.2016.10.005

Chen, Z., & Lux, T. (2018). Estimation of sentiment effects in financial markets: A simulated method of moments approach. *Computational Economics*, *52*(3), 711–744. https://doi.org/10.1007/s10614-016-9638-4

Corea, F. (2016). Big Data Analytics: A Management Perspective. In Studies in BigData. https://doi.org/10.1080/23270012.2015.1082449

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *Journal of Finance*, *66*(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Da, Z., Engelberg, J., & Gao, P. (2011a). In search of attention. *Journal of Finance*. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies*. https://doi.org/10.1093/rfs/hhu072

Daniel, M., Neves, R. F., & Horta, N. (2016). Company event popularity for financial markets using twitter and sentiment analysis. *Expert Systems with Applications*, *71*, 111–124. https://doi.org/10.1016/j.eswa.2016.11.022

D'Avanzo, E., Pilato, G., & Lytras, M. (2017). Using twitter sentiment and emotions analysis of Google trends for decisions making. *Program: Electronic Library and Information Systems*, *51*(3), 322–350. https://doi.org/10.1108/PROG-02-2016-0015

Dergiades, T., Milas, C., & Panagiotidis, T. (2015). Tweets, Google trends, and sovereign spreads in the GIIPS. *Oxford Economic Papers*, *67*(2), 406–432. https://doi.org/10.1093/oep/gpu046

Dzielinski, M. (2012). Measuring economic uncertainty and its impact on the stock market. *Finance Research Letters*, *9*(3), 167–175. https://doi.org/10.1016/j.frl.2011.10.003

Fernández-Villaverde, J., & Rubio-Ramírez, J. F. (2007). Estimating macroeconomic models: A likelihood approach. *Review of Economic Studies*, *74*(4), 1059–1087. https://doi.org/10.1111/j.1467-937X.2007.00437.x

Fisher, I. E., Garnsey, M. R., & Hughes, M. E. (2016). Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. *Intelligent Systems in Accounting, Finance and Management*, *23*(3), 157–214. https://doi.org/10.1002/isaf.1386

Franke, R., & Westerhoff, F. (2012). Structural stochastic volatility in asset pricing dynamics: Estimation and model contest. *Journal of Economic Dynamics and Control*, *36*(8), 1193–1211. https://doi.org/10.1016/j.jedc.2011.10.004

Ghonghadze J, Lux T. 2016. Bringing an elementary agent-based model to the data: Estimation via GMM and an application to forecasting of asset price volatility. *Journal of Empirical Finance 37*(C): 1–19. https://doi.org/10.1016/j.jempfin.2016.02.002

Goel, S., & Uzuner, O. (2016). Do sentiments matter in fraud detection? Estimating semantic orientation of annual reports. *Intelligent Systems in Accounting, Finance and Management*, *23*(3), 215–239. https://doi.org/10.1002/isaf.1392

Guéranger, D. (2009). Contradictions and conflicts in sociological writing: The rewriting of an interview by Pierre Bourdieu. *Social Science Information*, *48*(4), 615–629. https://doi.org/10.1177/0539018409344783

Guerini, M., & Moneta, A. (2017). A method for agent-based models validation. *Journal of Economic Dynamics and Control*, *82*, 125–141. https://doi.org/10.1016/j.jedc.2017.06.001

Jahangirian, M., Eldabi, T., Garg, L., Jun, G. T., Naseer, A., Patel, B., … Young, T. (2011). A rapid review method for extremely large corpora of literature: Applications to the domains of modelling, simulation, and management. *International Journal of Information Management*, *31*(3), 234–243. https://doi.org/10.1016/j.ijinfomgt.2010.07.004

Jang, T.-S. (2015). Identification of social interaction effects in financial data. *Computational Economics*. *45*, 207–238.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, *5*(1), 1–25. https://doi.org/10.2307/1390750

Kukacka, J., & Barunik, J. (2017). Estimation of financial agent-based models with simulated maximum likelihood. *Journal of Economic Dynamics and Control*, *85*, 21–45. https://doi.org/10.1016/j.jedc.2017.09.006

Lux, T. (2018). Estimation of agent-based models using sequential Monte Carlo methods. *Journal of Economic Dynamics and Control*. https://doi.org/10.1016/j.jedc.2018.01.021

Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: Mental imagery as mental emulation. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *364*(1521), 1273–1280. https://doi.org/10.1098/rstb.2008.0314

Muñoz-Expósito, M., Oviedo-García, M. Á., & Castellanos-Verdugo, M. (2017). How to measure engagement in twitter: Advancing a metric.

Internet Research, *27*(5), 1122–1148. https://doi.org/10.1108/IntR-06-2016-0170

Narita, F., & Yin, R. (2018). In search of information: Use of Google Trends' data to narrow information gaps for low-income developing countries. IMF working papers, working paper no. 18/286.

O'Leary, D. E. (2015). Twitter mining for discovery, prediction and causality: Applications and methodologies. *Intelligent Systems in Accounting, Finance and Management*, *22*(3), 227–247. https://doi.org/10.1002/isaf.1376

Oviedo-García, M. Á., Muñoz-Expósito, M., Castellanos-Verdugo, M., & Sancho-Mejías, M. (2014). Metric proposal for customer engagement in Facebook. *Journal of Research in Interactive Marketing*, *8*(4), 327–344. https://doi.org/10.1108/JRIM-05-2014-0028

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google trends. *Scientific Reports*, *3*, 1684. https://doi.org/10.1038/srep01684

Qiu, L. X., & Welch, I. (2004). Investor sentiment measures. Retrieved from https://doi.org/10.2139/ssrn.589641

Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking and Finance*, *84*, 25–40. https://doi.org/10.1016/j.jbankfin.2017.07.002

Riloff, E., & Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In *EMNLP '03 Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics: Stroudsburg, PA; 105–112. https://doi.org/10.3115/1119355.1119369

Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, *346*(6213), 1063–1064. https://doi.org/10.1126/science.346.6213.1063

Shen, D., Urquhart, A., & Wang, P. (2019). Does twitter predict Bitcoin? *Economics Letters*, *174*, 118–122. https://doi.org/10.1016/j.econlet.2018.11.007

Siganos, A., Vagenas-Nanos, E., & Verwijmeren, P. (2017). Divergence of sentiment and stock market trading. *Journal of Banking and Finance*, *78*, 130–141. https://doi.org/10.1016/j.jbankfin.2017.02.005

Stahl, D., & Hauth, J. (2011). PF-MPC: particle filter-model predictive control. *Systems and Control Letters*, *60*(8), 632–643. https://doi.org/10.1016/j.sysconle.2011.05.001

Stephens-Davidowitz, S. & Varian, H.(2015). A Hands-on Guide to Google Data. Google, Inc.

Strapparava, C., & Valitutti, A. (2004). Word Net-affect: An affective extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation, LREC 2004*.

Suppala, K., & Rao, N. (2019). Sentiment analysis using naïve Bayes classifier. *International Journal of Innovative Technology and Exploring Engineering*, *8*(8), 265–269. https://doi.org/10.5815/ijieeb.2016.04.07

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Tsukioka, Y., Yanagi, J., & Takada, T. (2018). Investor sentiment extracted from internet stock message boards and IPO puzzles. *International Review of Economics and Finance*, *56*, 205–217. https://doi.org/10.1016/j.iref.2017.10.025

Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, *55*(4), 919–926. https://doi.org/10.1016/j.dss.2012.12.028

Zhu, J. J. H., Wang, X., Qin, J., & Wu,L. 2012. Assessing public opinion trends based on user search queries: validity, reliability, and practicality. Paper presented at the *Annual Conference of the World Association for Public Opinion Research*, Hong Kong, 14–16 June.

How to cite this article: Mnif E, Jarboui A, Hassan MK, Mouakhar K. Big data tools for Islamic financial analysis. *Intell Sys Acc Fin Mgmt*. 2020;1–12. https://doi.org/10.1002/isaf.1463