cogent
economics
& finance

# FINANCIAL ECONOMICS | REVIEW ARTICLE

# Analysing stock market data—Market sentiment approach and its measures

K.K. Moseki[1]* and K.S. Madhava Rao[1]

*Corresponding author: K.K. Moseki, Department of Statistics, University of Botswana, Private Bag 0022, Gaborone, Botswana
E-mails: mosekikk@mopipi.ub.bw, mosekikk49@gmail.com

**Abstract:** This paper states that market sentiments are central to any financial data analysis. A vivid distinction is made between studying financial data in terms of the concept of volatility and in rapport to analysing financial data in terms of market sentiments. The former is an existing approach that is extensively used and the latter is a proposed tactic. Methods of devising constructs for defining relative and absolute market sentiments are also discussed. Patterns of market sentiments in terms of the model parameters are discussed and a few new measures that capture the hypothesized market sentiments are proposed. As an application of the proposed line of approach, this study analyses weekly market sentiments that govern Domestic Company Indices of Botswana Stock Exchange.

Subjects: Advanced Mathematics; Applied Mathematics

Keywords: divergence measures; equity price; market sentiments; multinomial model; volatility; weekly states

JEL classifications: C 19; G13; G 14

## 1. Introduction
The natural up and down patterns of the daily stock prices is indicative of the stochastic nature of the financial markets. Despite such price movements investors risk their funds in the stocks with the

## ABOUT THE AUTHORS
K.K. Moseki is a lecturer in the Department of Statistics, University of Botswana. He is pursuing his PhD in Statistics at the same department. He has published research papers in regional journals. His research interests include financial data analysis, time series, among others.

K.S. Madhava Rao is a professor in the Department of Statistics, University of Botswana. He has published extensively in national, regional and international peer-reviewed journals with high-impact factors. His research areas include nonparametric inference, empirical likelihood, social sciences, agricultural and financial data analysis.

## PUBLIC INTEREST STATEMENT
Analysis of market sentiments is central to any financial data analysis. Rao and Moseki (2009) advocated market sentiments approach to analyse financial data as opposed to the time series based GARCH model approach which has been well researched over the past several decades. In their work, a clear distinction was made between the existing approach of studying financial data in terms of extensively used concept of volatility and the proposed approach in terms of market sentiments. In this paper, we further explore the types of market sentiments and device constructs for defining relative and absolute market sentiments. A few possible patterns of market sentiments in terms of the model parameters are discussed and new measures which capture the hypothesized market sentiments are proposed. As an application of the proposed line of approach, we analyse weekly absolute market sentiments that govern Domestic Company Indices of Botswana Stock Exchange.

expectation of making more future returns. The variation in the stock returns due to daily price changes is generally termed as volatility that is measured by the standard deviation of the returns. Investors in stock market are interested in the volatility of stock prices as high volatility could be interpreted as huge losses or profits at the expense of greater uncertainty (Gujarati, 2003). Volatility is something good but that can only be bad if the stock price swings are unusually very high and rapid over short time periods as it makes financial planning difficult (Islam, 2013). This means that high fluctuations in stock prices can possibly increase both risk and uncertainty about the future returns. Since prediction of stock is part and parcel of stock trading, if the market performance is unstable then investors will be clueless as what the future holds for them and hence increased uncertainty about future price changes. As a result of uncertainty investors can fear to take risk and continue to fund their investments.

A volatile market inhibits individual retailers and companies to raise funds in the capital markets. Islam (2013) cited that uncertainty causes loss of investor confidence that is essential in trading of stocks particularly in encouraging further investment and leverage decision. He further mentions that uncertainty can aggravate volatility. The knowledge of volatility is of crucial importance in many areas of finance in enabling investors to manage risks and optimize profits (Hull, 2003). The fluctuations in the asset prices are widely believed to be the cause of changes in the economic and geo-political environmental factors such as stocks, bonds, exchange rates, tradable options and banking instruments among others. The perceived knowledge of the aforementioned contributing factors along with analysis of data gives a player in any stock market a strong and reliable tool to trade.

The financial data analysis is focused on the prediction of the future prices of security and as such they are model-based strategies. These are mainly tailored for the needs of well-informed market players that include individual investors, financial institutions and mutual fund managers. It is widely believed that, security prices exhibit considerable variations from time to time and these variations can be objectively captured by well-defined market volatility phenomena. The concept of volatility is imperative in the analysis of financial time series data. Volatility is the unpredictable change of a variable over time. In the context of finance and economics, volatility refers to the unstable states of economic and financial variables, these variables include stock prices, exchange rates, inflation rates, volume of trade and currency returns. There are several definitions of volatility and the common denominator is variability. The different definitions of volatility are distinct in how they discuss the measurement of this variability. The causes of volatility could be ascribed to several factors; such as local and global economic environments, maturity and delivery date of instruments, changes in trade volume, practices or patterns and new information among others (Hull, 2012).

### 1.1. Objectives and implications of the study

A common assumption in applying GARCH-type models to financial data is that the conditional return series is conditionally normally distributed. In practice, however, the shape of the daily distribution of return series is much broader than the commonly assumed Gaussian distribution. For example, truncated Levy distribution, Student's distribution, and log normal distribution among others have been used. It is pertinent to indicate that in this paper, we do not assume any specific form of the probability distribution for stock return series. Thus, we adopt a nonparametric approach to pursue the objectives of our study.

The central contribution of this paper is that it suggests an alternative approach for modelling and related analysis of stock returns. In the new approach suggested here, volatility in weekly means of stock returns is conceived to be a function of the mean and its standard error and the perceived volatility is classified into various market sentiment states. Thus, a fairly accurate knowledge of such states of volatility in stock returns will be of great help to a prospective or an existing market player. The main objectives of the paper are as follows:

(1) To appreciate the distinction between absolute and relative market sentiments that govern stock return data,

(2) To devise suitable data-driven rules to classify volatility in stock returns into market sentiment states,

(3) To build a probability model for capturing volatility in stock returns in terms of different market sentiment states,

(4) To formulate possible hypotheses about the states of market sentiments,

(5) To propose different measures to capture the underlying patterns those exist in states of the market sentiments and also to test their significance.

The implications of the current work are many folds. A few among these are stated below:

(1) The results in this paper indicate an alternative approach to analyse financial data, namely the market sentiment approach, as opposed to the volatility approach that is in vogue,

(2) It is established that the short term movements in a financial data can be analysed in terms of relative market sentiments, while the analysis of absolute market sentiments provide indicators of long-term movements in financial return series,

(3) Any decision-making that an investor takes on his investment portfolio is heavily dependent on the direction of the market sentiments. Basically, financial decisions are generally based upon the trade-off between risk and return, it is therefore imperative for investors to make proper decisions in order to avoid too much loss of business. In this paper, we are presenting measures that will assist in decision-making and future planning of investment portfolio,

(4) Due to the novelty of the proposed approach, it is hoped that the line of analysis suggested here will be of interest to both academicians and stock market players.

The paper is structured as follows: After the introduction, second section briefly discusses literature review. The third section outlines methodological framework. In the fourth section, as an application of the proposed model, we analyse Botswana Stock Exchange data from 1999 to 2013, and presents the results and discussions. The probability models are based on the weekly indices. The last section provides conclusions and recommendations, followed by a list of references used in this study.

## 2. Literature review

The literature on economic and financial variable data analysis is indeed quite extensive. Forecasts of the distribution of such variables like asset returns are of great importance in the context of financial risk management. In pricing options, anticipated volatility over the life of the option is the crucial unknown parameter. One of the most popular tools used to explain volatility clustering in option pricing is the Autoregressive Conditional Heteroscedasticity (ARCH) class of models which was invented by Engle (1982). The theoretical frame work of modelling volatility advocated by Engle captures the time variability of the variance of returns by imposing an autoregressive structure on the conditional second moment of returns. Over the years there have been a number of extensions to the basic ARCH model. The most important of these is the generalised ARCH (or GARCH) model developed by Bollerslev (1987) which includes lagged values of the conditional variance as extra regressors. Subsequently, many other ARCH-type processes have been developed to capture various dynamics. For additional references and expository studies using ARCH type of models and also for exhaustive theoretical treatment of volatility models, we refer to Blake (2005) and Sheldon (1994).

Generally, market sentiments are perceived to be either absolute or relative. Absolute market sentiments can be defined with reference to a base period greater than a month such as a quarter or bi-annual or annual. In this case, market sentiments are constructed based on the entirety of data corresponding to the base period. For example, the weekly market sentiments can be constructed based on the daily stock prices of a given year. It may be pointed out that absolute market sentiments are useful to compare the movements in market sentiments over blocks of years and hence

may be of use to players with long-term interest. In other words, absolute market sentiment captures the long-term impact on volatility. Relative market sentiments are defined with reference to shorter base period such a week or month. In this case, market sentiments are more dynamic in the sense that constructs compare weekly or monthly data from time to time. Relative market sentiments are seen to be useful to an active player with short-term interest. A large share of overall movement of an individual stock has been attributed to market sentiments. Forecasting market sentiments in financial data such as stock indices, volume of trade is a formidable area of research in financial economics. There are innumerable causes, as mentioned earlier, that propel market sentiments in a financial data. Classically, market sentiment is monitored by a variety of technical and statistical methods such as the number of advancing vs. declining stocks and new highs vs. new lows comparisons.

There are multiple of studies dealing with stock return volatility from an investor's perceptive. In most of these studies, researchers preferred to use the terminology investor sentiment rather than the "market sentiment". For the sake of completeness, we outline a few of these, keeping in mind that none of these studies follow the approach that is advocated in this paper.

Within the scope of behavioural finance, many studies have gone on to analyse factors relating to the mentality of investors, chief among others being volatility. A comprehensive review of the prior studies based on the role of market sentiments in the financial markets was investigated by Shiller (1999) who examined the behaviour of investors in the financial markets; Hirshleifer and Shumway (2003) demonstrated, from the psychological point of view, that since investors have acute shortage of information, investor sentiment is a factor for consideration in the formulation of any investment strategy. In other words, when investors lack complete information regarding their investment, they are believed to make their choices based on unreasonable behaviour and personal considerations.

Lee, Jiang, and Indro (2002) investigated the relationship between volatility, stock returns and investor sentiment. They estimated a GARCH-in-mean model which includes contemporaneous shifts in investor sentiment in the mean equation and lagged shifts in investor sentiment in the conditional volatility variance equation. They used the survey indicators provided by investors' intelligence to examine the impact of changes in investor sentiment on the conditional volatilities and indices. They found that bullish (bearish) changes in investor sentiment result in downward (upward) adjustments in volatility.

Wang, Keswani, and Taylor (2006) and Lee and Kumar (2006) showed that the index of investor sentiment has no predictive power in the stock market return, or at best very limited ability. This has not diluted the appeal by the majority of studies. Within the prior studies by Brown (1999), Fisher and Statman (2000), Lee et al. (2002), Baker and Wurgler (2004), Baker and Stein (2004), Brown and Cliff (2004) and Lee and Kumar (2006) tend to agree that there is correlation between the investor sentiment index and stock market returns, and this could be used to predict the index of return within the stock market. Yoshinaga and de Castro (2012) estimated a pricing model that relates stock rate of return to market sentiment index and other factors and the results confirm that the sentiment variable plays a relevant role.

Rao and Moseki (2009) analysed the data on Domestic Company Indices of Botswana Stock Exchange using a probabilistic approach to market sentiments. The methodology developed in that paper is applied to weekly and monthly DCI of the Botswana Stock Exchange. The findings in that study are; the weekly and monthly DCI of the BSE are volatile; the market environment in BSE is investor-friendly as the bubble sentiments are the modal sentiments; the variation in monthly DCI is more stable than the weekly DCI; the weekly trading is more speculative than the monthly trading. Finally, from the seller's point of view, returns from investment are more if the decision to sell scrip is based on weekly volatility analysis rather than monthly volatility analysis. Rao and Moseki (2011) provided an alternative approach, namely a Markovian approach to the study of market sentiments. An application of their methodology to the weekly DCI revealed that "greedy" market sentiment

cogent •• economics & finance

prevailed during the study period as the long-term probability related to the "greedy" sentiment was high and close to 0.5. Rao and Ramachandran (2014) used the methodologies advocated by Rao and Moseki (2009) to analyse exchange rate market sentiments of six major global currencies against the Botswana Pula. In this research, the authors considered relative market sentiments. The study has generated the following results; weekly and monthly exchange rate market sentiments appear to fall into two regimes of symmetric and asymmetric sentiments; all the six currencies exhibited smiling sentiments giving the scenario of asymmetric market sentiments; all the currencies have clusters of sentiment states which are strongly persistent revealing that the exchange rate trading across the currencies and the years are buoyant. In conclusion, the literatures reviewed do not capture the direction of the market sentiments which is a key variable of interest to the traders in the market whenever they want to make investment decisions. An attempt is also made to fill this gap by devising suitable measures.

## 3. Methodology

### 3.1. Market sentiment constructs

Because this paper focuses on financial applications, we will use financial notation. Let "$t$" refer to the trading day and "$N$" the total number of trading days in a study period. Further, let "$w$" denote the week, so that $W$ corresponds to the total number of weeks in a study period. Let the number of trading days in a week be denoted by $n_w$. In a typical stock market, $n_w$ could vary between 1 and 5. Let for a week $w = 1, ..., W, P_t^w$ denote the equity index of day "$t$", $t = 1, ..., n_w$, $w = 1, ..., W$. Our analysis will be based on the log returns of stock index (DCI) defined by

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right), \quad t = 2, 3, \ldots, n_w$$

For analysing weekly DCI returns, for a week $w = 1, ..., W$, we define the average returns by

$$\bar{r}_w = \frac{1}{n_w} \sum_{t=1}^{n_w} r_t, \quad w = 1, \ldots, W, \tag{1}$$

and the sample variance of equity return index for the week $w$ by

$$s_w^2 = \frac{1}{(n_w - 1)} \sum_{t=1}^{n_w} \left(r_t - \bar{r}_w\right)^2. \tag{2}$$

Similarly, for the entire period $N$, we have the combined mean and the combined sample variance of the equity returns, respectively, given by,

$$\bar{r}_c = \frac{\sum_{w=1}^{W} \sum_{t=1}^{n_w} r_t}{N} = \frac{\sum_{w=1}^{W} n_w \bar{r}_w}{\sum_{w=1}^{W} n_w}, \tag{3}$$

$$S_c^2 = \frac{1}{W} \sum_{w=1}^{W} \sum_{t=1}^{n_w} \left(r_t - \bar{r}_w\right)^2 = \frac{1}{W} \sum_{w=1}^{W} (n_w - 1) s_w^2. \tag{4}$$

If we assume that $r_t$'s come from a population with finite mean $\mu$ and finite variance $\sigma^2$, then we know that $E(\bar{r}_w) = \mu$, $\text{Var}(\bar{r}_w) = \frac{\sigma^2}{n_w}$ and let $Z = \frac{(\bar{r}_w - \mu)}{\sigma} \sqrt{n_w}$. We shall use the statistic $Z$ to define classification rules for the market sentiments.

As in Rao and Moseki (2009), Rao and Ramachandran (2014), we consider the following seven type of market sentiments.

The market sentiments, from the standpoint of a seller in a stock market can be seen to belong to the following categories.

$S_1$: *The depressed/fear sentiment*. This is a state in which the equity index witnesses a sharp fall, enough to cause panic to a seller holding stocks. This may lead to distress selling and shatter investor confidence. This sentiment may also witness eventual bankruptcy or consolidation or regulatory intervention. Further, raising funds from the capital market may be extremely difficult.

$S_2$: *The bearish sentiment*. This is a state in which the equity index witnesses a significant downward trend, enough to cause frustration to a seller holding stocks. This may lead to an offloading exercise of stocks and undermine investor confidence. Moreover, raising funds from the capital market may be adversely affected.

$S_3$: *The downward sentiment*. This is a state in which the equity index witnesses a gradual fall enough to sound a warning to an investor. This may lead to an offloading exercise of stocks and may dent investor confidence. Sustained attempts may be needed to raise funds from the capital market.

$S_4$: *The tranquil sentiment*. This is a state in which the equity index witnesses normal swings. The investor continues to be an active player in the stock market. This state may also be called "range bound" or "uncertain".

$S_5$: *The upward sentiment*. This is a state in which the equity index witnesses a gradual upward swing enough to cause a feeling of encouragement to an investor. The environment for fund raising may be encouraging.

$S_6$: *The bullish sentiment*. This is a state in which the equity index witnesses a significant upward swing; propelling speculation by a seller holding stocks. Further, raising funds from the capital market may be easy.

$S_7$: *The greed/bubble sentiment*. This is a state in which the equity index witnesses a sharp rise, enough to prompt a seller either to dispose of his load or to hold on in the hope of further appreciation leading to a bubble like condition. This state assures the seller of massive gains for having waited a while. In addition to this, raising funds from the capital market may be spontaneous.

In order to reflect these seven market sentiments, we associate the following classification rules, based on the statistic $Z$, with the states $S_j, j = 1, 2, 3, 4, 5, 6, 7$. Let $\hat{\mu}$ and $\hat{\sigma}^2$ be consistent estimators of $\mu$ and $\sigma^2$, respectively.

Then for a week $(w + 1), w = 1, \ldots, (W - 1)$, the equity index is said to belong to

(i) State $S_1$, if $\bar{r}_{w+1} < \hat{\mu} - 3 \dfrac{\hat{\sigma}}{\sqrt{n_w}}$

(ii) State $S_2$, if $\hat{\mu} - 3 \dfrac{\hat{\sigma}}{\sqrt{n_w}} \leq \bar{r}_{w+1} < \hat{\mu} - 2 \dfrac{\hat{\sigma}}{\sqrt{n_w}}$

(iii) State $S_3$, if $\hat{\mu} - 2 \dfrac{\hat{\sigma}}{\sqrt{n_w}} \leq \bar{r}_{w+1} < \hat{\mu} - \dfrac{\hat{\sigma}}{\sqrt{n_w}}$

(iv) State $S_4$, if $\hat{\mu} - \dfrac{\hat{\sigma}}{\sqrt{n_w}} \leq \bar{r}_{w+1} < \hat{\mu} + \dfrac{\hat{\sigma}}{\sqrt{n_w}}$

(v) State $S_5$, if $\hat{\mu} + \dfrac{\hat{\sigma}}{\sqrt{n_w}} \leq \bar{r}_{w+1} < \hat{\mu} + 2 \dfrac{\hat{\sigma}}{\sqrt{n_w}}$

(vi) State $S_6$, if $\hat{\mu} + 2 \dfrac{\hat{\sigma}}{\sqrt{n_w}} \leq \bar{r}_{w+1} \leq \hat{\mu} + 3 \dfrac{\hat{\sigma}}{\sqrt{n_w}}$

(vii) State $S_7$, if $\bar{r}_{w+1} > \hat{\mu} + 3\dfrac{\hat{\sigma}}{\sqrt{n_w}}$

$$(5)$$

The value of the multiplier of $\dfrac{\hat{\sigma}}{\sqrt{n_w}}$ between −3 and 3 are so chosen as the cut off points of the standard normal distribution that in essence reflects a symmetric distribution. These values reflect normal to extreme shift on either side of the middle. So, the whole idea is to try to create symmetric market sentiments of the financial stock market data. The notion of symmetric distribution of asset returns to most investors is an ideal environment in the sense that changes in the security return is the same from either side of the middle. So, multipliers used are equivalent to the action and warning limits of the statistical process control (See, for example, Grant & Leavenworth, 1980).

In the approach that we are proposing, we use these values to generate categories of volatility of asset returns we referred as market sentiments. The seven-states market sentiments constructs is general, in a sense that they are capable of reflecting most of the market sentiments usually observed in a vibrant stock market economy. In a less vibrant stock market economy, one may continue to work with smaller set of states. But if the market sentiments appear to be uniform, say, four categories, one may be inspired to use the states defined on the basis of quartiles, and the multipliers of the standard error may correspond to the quartiles of the standard normal distribution. These multipliers will be used to construct four market sentiment states.

Suppose we take $\hat{\mu} = \bar{r}_w$, the weekly average of log returns of equity prices as an estimator of $\mu$, and $\hat{\sigma}^2 = s_w^2$, the weekly sample variance of log returns as an estimator of $\sigma^2$. Then, the classification rule (5) with these estimators will define relative market sentiments as considered in Rao and Moseki (2009). On the other hand, if we use the combined estimators $\hat{\mu} = \bar{r}_c$, $\hat{\sigma}^2 = S_c^2$, and replacing $n_w = W$, in (5), the classification rules generate absolute market sentiments.

### 3.2. Modelling market sentiments

Following the assumptions and the notations of Rao and Moseki (2009), in general, the vector of counts of market sentiment states $S = (S_1, \ldots, S_k)$ follows a multinomial distribution with the joint probability mass function (p.m.f.) given by

$$p(f_1, f_2, \ldots, f_k) = \frac{W!}{f_1! f_2! \cdots f_k!} \theta_1^{f_1} \theta_2^{f_2} \ldots \theta_k^{f_k},$$

$$(6)$$

for all $0 \leq f_j \leq W$, $\sum_{j=1}^{k} f_j = W$ and $\sum_{j=1}^{k} \theta_j = 1$; where, $f_j$ denotes the number of occurrences of the state

$S_j$ in an independent sequence of the phenomena observed for $W$ weeks and $\theta_j = P_r(\bar{r}_{w+1} \in S_j)$, $j = 1, 2, \ldots, k$. In general, the parameters $\theta_j$'s in the model given by (6) are unknown and can be estimated by their empirical estimates

$$\hat{\theta}_j = \frac{f_j}{W}, \quad j = 1, 2, \ldots, k.$$

$$(7)$$

### 3.3. Hypotheses about market sentiments

The market sentiments witness different types of patterns with respect to time, namely bullish pattern (increasing); bearish pattern (decreasing); symmetric pattern criss-crossing (evenly around a calm sentiment). While a bullish or a bearish run may be imperative consequence to an economic or geo-political vital event and therefore may be of short duration, a symmetric pattern is perceived to be the rule of vibrant market economies. These patterns can be effectively summarized in the terms of the model parameters as follows:

$$H_I : \theta_1 \leq \ldots \leq \theta_k \tag{8}$$

$$H_D : \theta_1 \geq \ldots \geq \theta_k \tag{9}$$

$$H_S : \theta_i = \theta_{k-i+1}, \quad i = 1, \ldots, k. \tag{10}$$

The pattern in (8) is indicative of an ordered increasing sentiment (Bull Run), while (9) refers to that of an ordered decreasing sentiment (Bear Run). The pattern in (10) reflects a symmetric pattern.

As an illustration for the sentiment pattern $H_1$, an *apriori* knowledge about a bull run may be characterized by $\theta_i = \frac{2i}{k(k+1)}, i = 1, 2, \ldots, k$. On the other hand, a bear run exhibited by $H_D$ will have $\theta_i = \frac{2(k-i+1)}{k(k+1)}, i = 1, 2, \ldots, k$.

When $\theta_i$'s are unknown and follow the sentiment pattern $H_s$ given by (10) an appropriate symmetric multinomial distribution model with $k = 7$, using the classification rule (5) is given by,

$$p(f_1, \ldots, f_7) = \frac{W!}{f_1! f_2! f_3! f_4! f_5! f_6! f_7!} \theta_1^{f_1+f_7} \theta_2^{f_2+f_6} \theta_3^{f_3+f_5} \theta_4^{f_4}, \tag{11}$$

for all $0 < f_j < W, \sum_{j=1}^{7} f_j = W$ and $\theta_4 = 1 - 2(\theta_1 + \theta_2 + \theta_3)$.

### 3.4. Measures of market sentiments

The seven-states symmetric multinomial distribution model was used by Rao and Moseki (2009) and Rao and Ramachandran (2014) to estimate measures of market sentiments and also to test hypothesis $H_s$, namely the symmetry of market sentiments of equity indices. Their approach was intuitively based on the classical chi-square $\left(\chi^2\right)$ measure of discrepancy between the observed and the expected counts and likelihood ratio principle ($\lambda$). The two measures considered by Rao and Moseki (2009), were given by

$$X^2(f) = \sum_{j=1}^{3} \frac{\left(f_j - f_{8-j}\right)^2}{f_j + f_{8-j}}, \tag{12}$$

$$\lambda(f) = \prod_{j=1}^{4} \left\{ \frac{\left(\frac{f_j + f_{8-j}}{2}\right)^{f_j + f_{8-j}}}{\left(f_j\right)^{f_j} \times \left(f_{8-j}\right)^{f_{8-j}}} \right\} \tag{13}$$

It can be easily shown that these discrepancy measures are related to certain members in the family of power divergence statistics proposed by Cressie and Read (1984).

These authors carried out test for symmetry of the market sentiment states for the asset returns of Domestic Company Indices in Botswana and exchange rate of market sentiments of Botswana Pula against six global currencies, respectively. For Rao and Moseki (2009), the test concluded that the weekly and monthly equity index data were highly asymmetric, while for Rao and Ramachandran (2014), it was revealed that two regime paths emerged, one for symmetry and the other for asymmetry.

Here below, we advance a few new measures of market sentiments that can be used for the types of market sentiments hypothesised in (8)–(10) above.

The measures given by (12) and (13) have the inherent disadvantage that they do not indicate the direction of market sentiments. Firstly, for the pattern $H_s$ given by (10), a variant of the measure (12) can be defined as follows:

$$\delta(f) = \sum_{j=1}^{3} \text{sgn}\left(f_{8-j} - f_j\right) \frac{\left(f_{8-j} - f_j\right)^2}{\left(f_{8-j} + f_j\right)}, \tag{14}$$

where sgn(.) is the sign function defined by $\text{sgn}(u) = 1$, if $u > 0$, $\text{sgn}(u) = 0$, if $u = 0$, $\text{sgn}(u) = -1$, else. This measure of market sentiments takes into accounts both the magnitude and direction of market sentiments, therefore especially appropriate when the market sentiments are asymmetric. For example, a large positive $\delta(f)$ would indicate smiling sentiment states, while a large negative $\delta(f)$ would indicate frowning sentiment. We observe that $\delta(f)$ is a measure based on the signed dispersion.

### 3.4.1. A measure of ideal market sentiments

A second measure of market sentiment is surmised on the theory that in an ideal market condition, more often, the sentiments should be concentrated around the calm sentiment state, namely the state $S_4$, when $k = 7$. An ideal market sentiment results in least speculations and minimal loss or gain to the player. As a measure of ideal market sentiments, the study proposes a measure which is akin to the mean sum of squares of deviations of the sentiment states taken from the calm sentiment state. Let $W$ denote the number of weeks in a year and $t_w$, $w = 1, ..., W$, be the value of the sentiment state corresponding to a week $w$. Note that $t_w$ will take one of the values from 1 to 7. For example, if $t_1 = 2$; then for the first week, the sentiment state is "2", if $t_2 = 6$; then for the second week, the sentiment state is "6" and so on. A dispersion measure of an ideal market sentiment is given by;

$$D(t_w) = \frac{1}{W} \sum_{w=1}^{W} (t_w - 4)^2. \tag{15}$$

It is evident that in an ideal market condition, $t_w = 4$ for all $w$ and therefore $D(t_w) = 0$. On the other hand, if the sentiments swing on either side of the normal threshold state $S_4$, $D(t_w)$ will be significantly different from 0. In the two extreme of cases; $D(t_w) = 9$. Thus, greater the values of $D(t_w)$, more dispersed are the weekly sentiments from the calm sentiment state $S_4$. In other words, $D(t_w)$ defines risky sentiments. In our context large value of $D(t_w)$ closer to nine would mean risky market sentiments in the sense that there are appreciable swings in the market sentiment states. It is pertinent to note that the measure $D(t_w)$ is an absolute measure of swings, but it fails to measure symmetry of market sentiments like the measures $\chi^2(f)$, $\lambda(f)$ and $\delta(f)$.

### 3.4.2. A measure of uncertainty of return series

Finally, uncertainties associated with market sentiment models given by (6) and (11) can also discussed in terms of information theoretic entropy. For example, we may use Shannon's entropy, as in Shannon (1948), for the model (6), which is given by;

$$\varepsilon_k(\theta) = - \sum_{i=1}^{k} \theta_i \log(\theta_i). \tag{16}$$

The choice of a logarithmic base corresponds to the choice of a unit for measuring information. Shannon himself discussed logarithm bases 2, 10 and e, but in practice entropy is always calculated using base 2 logarithm. Henceforth, this paper uses base 2 in computations. When $\theta_i$'s are unknown, we compute $\varepsilon_k(\theta) = - \sum_{i=1}^{k} \theta_i \log(\theta_i)$, using the empirical estimates of $\theta_i$'s given by (7). On the other hand, when $\theta_i$'s are perceived to be known *apriori*, we may use the formula given by (16) directly. The most rational choice of $\theta_i$'s is $\theta_i = k^{-1}$, $i = 1, ..., k$, in which case the maximum entropy will be equal to $\log(k)$. Thus, for $k = 7$, $\varepsilon_7(\theta) = \log(7) = 2.807$, for an ideal band of market sentiments. If one uses the classification rule (5) stated in Section 3.1, using the empirical rule of large sample approximation, *apriori*, we have, $\theta_1 = 0.00135 = \theta_7$, $\theta_2 = 0.0215 = \theta_6$, $\theta_3 = 0.1359 = \theta_5$, $\theta_4 = 0.6826$. In this case, $\varepsilon_7(\theta) = 1.423$. This paper observes that the classification rule with these *apriori* probabilities provides the most extreme band of market sentiments. Therefore, it may be argued that model uncertainties as defined by Shannon's entropy will vary in between the entropy measure corresponding to the extreme ($H_I$ or $H_D$) pattern and the symmetric ($H_s$) band of market sentiments. Thus, under this assumption we have; $1.423 \leq \varepsilon_7(\theta) \leq 2.807$ and this interval may be used to ascertain the

significance of any computed entropy for the observed market sentiment counts for the given year. Similarly, with the priori choice of model parameters given by

$$H_I : \theta_i = \frac{2i}{k(k+1)} \quad \text{and} \quad H_D : \theta_i = \left( \frac{2(k - i + 1)}{k(k+1)} \right), \quad i = 1, \dots, k,$$

Using (16), it can be shown that the entropy value for the market sentiment scenario given by (8) and (9) is given by $\varepsilon_7(\theta) = 2.61$.

### 3.4.3. Divergence measures

Alternatively, one may use the generalized directed divergence measure proposed by Cressie and Read (1984) or Bregman (1967) divergence. For the most rational choice of $\theta_i$'s, namely $\theta_i = k^{-1}, i = 1, \dots, k$, it can be shown that a certain member of Cressie–Read divergence measure is given by

$$I_k(f) = \sum_{i=1}^{k} \frac{f_i}{W} \ell n \left( \frac{k f_i}{W} \right). \tag{17}$$

Further, it can be shown that the Bregman divergence based on the squared Euclidean distance is given by

$$d_\phi(f) = \sum_{i=1}^{k} \frac{f_i^2}{W^2} - \frac{1}{k}. \tag{18}$$

It is easy to show that $I_k(f) = 0, d_\phi(f) = 0$ when the observed frequency of market sentiment states are in perfect synchrony with the rational choice of $\theta_i$'s.

### 3.5. Data and basic statistics

The daily stock price index data used in this paper are the daily Domestic Company Indices (DCI) collected from the Botswana Stock Exchange over the period from 11 January 1999 to 31 December 2013. The daily index returns are expressed in the continuously compounded returns calculated as $r_t = \log(p_t) - \log(p_{t-1})$, where $p_t$ and $p_{t-1}$ are the index prices on day $t$ and $t - 1$ respectively. The market sentiments are calculated for the weeks in the calendar years 1999–2013, assuming that each starting and ending week in a calendar year has at least two trading days. The movement in DCI is classified into the seven-states $S_1, S_2, \dots, S_7$ based on the classification rule described by (5). The data can be downloaded from https://www.bse.co.bw or obtained from the first author. The details of calculations of weekly averages, standard deviations and states to which each weekly DCI belongs to are not reported in this paper but may be obtained from the first author on request. Before proceeding further for formal analysis of market sentiment states, we plot the weekly return series and their volatility over the study period as shown in Figure 1 in order to understand the likely nature of the return series.

It can be observed that the graphs displayed Figure 1 exhibit considerable swings or volatility in the return series over the sample period. The spikes in the return plots are the graphical evidence implying that the volatility is time varying conditional variance. It also indicates the presence of volatility clustering effect in the return series of DCI, that is to say, the series exhibit some periods of high volatility followed by other periods of high volatility and some periods of low volatility followed by other periods of relatively low volatility. Volatility clustering is a sign of positive autocorrelation in the squared returns.

The results from the descriptive statistics reported in Table 1 show that during the period of study, the Domestic Company Indices (DCI) of Botswana stock market observed the weekly mean return of

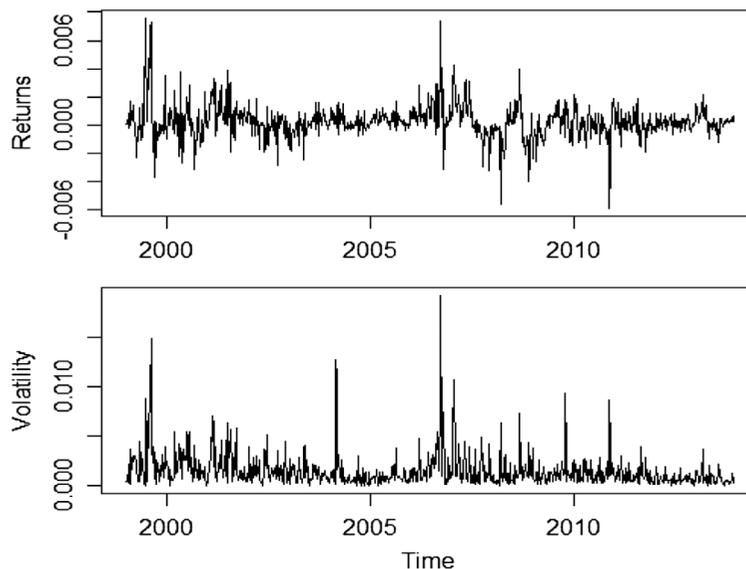**Figure 1. Weekly returns and volatility returns from January 1999 to December 2013.**



| Table 1. Summary statistics | |
|---|---|
| **Statistics** | **DCI return indices** |
| Mean | 0.00027 |
| Maximum | 0.16010 |
| Minimum | −0.16250 |
| Std. Dev | 0.00122 |
| Skewness | 0.90772 |
| Kurtosis | 7.80332 |
| Jarque–Bera (JB) | 2095.3 |
| Probability J.B | 0.0000 |
| ADF-test statistics | −7.1115 |
| Probability | 0.01 |
| No. of observations | 774 |

0.00027. The corresponding volatility of the DCI measured by standard deviation is 0.00122. The high volatility offers the possibility of high rate of returns. The return series for DCI shows positive skewness suggesting that the distribution has a long right tail meaning there are some extreme return values in the right tail of the series. The excess value greater than 3 for kurtosis indicates fat tail characteristics of the security returns distribution. The Jarque–Bera (JB) test of normality statistically and significantly rejects the null hypothesis of normality as indicated by the smallest significance level that is below 0.01. The test suggests that the distribution of the return series is non-normal.

### 3.5.1. Data stationarity test (unit root test)

In order to assess whether the financial time series (returns) are stationary or not the researchers have applied the augmented Dickey–Fuller (ADF) test. The results show that the ADF test statistics rejected the null hypothesis of the existence of the unit root in the return series as indicated by the highly significant value (<0.0001) of the associated probability of the ADF test statistic. The test summarizes that the return series of DCI is stationary and therefore the application of the stochastic models to return series is appropriate to examine the dynamic behaviour of volatility over time.

cogent ·· economics & finance

## 4. Results and discussion

Table 2 presents the summary of absolute market sentiments of DCI of BSE for the year 1999–2013. The multinomial model is highly skewed towards the two extreme sentiments (frowning and smiling sentiments). This is an illustration of a discrete bathtub type of a multinomial model. The practical significance of this finding is that the stock market in Botswana is characterized by extreme bulls and bears that present difficulties for potential investors willing to do business as a trader.

Table 3 presents the five measures of weekly absolute market sentiments. It is to be noted that missing values indicate that the corresponding measure is not defined due to zero counts. Based on the measure, $\chi^2(f)$, that takes values between 0 and infinity, the empirical observation reveals that the weekly stock return market sentiments more or less appear to fall in two regimes of symmetric and asymmetric sentiments. The findings point to 1999 and 2004 through 2006 as asymmetric, otherwise the rest of the years seem to be symmetric. The summary findings on $\delta(f)$ whose values range between $(-\infty, \infty)$, where intuitively small values would indicate symmetry of the market sentiments, indicates that the DCI market experienced periods of bear performance and periods of bull performance for study time periods. The measure reveals that for the study period considered the market movement was more towards the bearish sentiments than the bullish one. Considering the measure $\lambda(f)$, there is evidence that stock market is branded by symmetric market sentiments for values of measure close to zero and are very small except for years where the measure could not be computed due to zero counts of sentiment states. In other words, lack of vibrancy in the DCI trading resulted in a number of zero counts that caused the measure untenable for some of the years. The other measure that was considered is the $D(t_w)$ that measures the dispersion of the market sentiments from the calm sentiment, that communicates market swings of sentiment states. The larger the values of the measure close to nine, the riskier it is to invest in equity market. There is empirical evidence that across all the study period there has been no considerable swings experienced in the stock market. On using the measure $\varepsilon_7(\theta)$, that detects uncertainty or the presence of either increasing or decreasing patterns of the market sentiments over the years considered for the study. From the analysis we may conclude that market sentiments during the years 1999–2013 do not reflect the pattern $H_I$ or $H_D$, since the computed $\varepsilon_7(\theta)'s$ in column (6) of Table 3 are significantly different from the reference value $\varepsilon_7(\theta) = 2.61$. However, since the computed values are somewhat nearer to the reference value $\varepsilon_7(\theta) = 1.423$, we may conclude that the market witnessed more or less symmetric pattern $H_S$ during the entire study period. The measure $I_k(f)$ is computable only for some years and

| Table 2. Weekly absolute market sentiments | | | | | | | |
|---|---|---|---|---|---|---|---|
| **States(weekly counts)** | | | | | | | |
| **Year** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| 1999 | 19 | 8 | 7 | 2 | 1 | 2 | 11 |
| 2000 | 18 | 3 | 1 | 7 | 3 | 3 | 17 |
| 2001 | 17 | 6 | 3 | 10 | 0 | 1 | 16 |
| 2002 | 13 | 0 | 3 | 13 | 5 | 1 | 18 |
| 2003 | 16 | 1 | 2 | 7 | 4 | 5 | 18 |
| 2004 | 11 | 6 | 11 | 9 | 1 | 5 | 10 |
| 2005 | 16 | 6 | 6 | 9 | 1 | 0 | 14 |
| 2006 | 14 | 7 | 7 | 9 | 3 | 0 | 12 |
| 2007 | 20 | 3 | 2 | 7 | 2 | 4 | 15 |
| 2008 | 17 | 2 | 1 | 8 | 3 | 4 | 18 |
| 2009 | 20 | 1 | 3 | 1 | 6 | 0 | 22 |
| 2010 | 13 | 2 | 2 | 12 | 2 | 3 | 19 |
| 2011 | 18 | 1 | 7 | 8 | 2 | 0 | 16 |
| 2012 | 18 | 6 | 2 | 3 | 3 | 3 | 18 |
| 2013 | 17 | 6 | 3 | 8 | 3 | 3 | 13 |

cogent ·· economics & finance

| Table 3. Measures of market sentiments | | | | | | | |
|--------|--------|--------|--------|--------|--------|--------|--------|
| **Year** | **Different market sentiment measures** | | | | | | |
| | $\chi^2$ (f) | $\lambda$ (f) | $\delta$ (f) | $D(t_w)$ | $\varepsilon_k(\theta)$ | $I_k$ (f) | $d_\phi$ (f) |
| 1999 | 10.23 | 0.00 | −10.2 | 6.36 | 0.70 | 0.39 | 0.10 |
| 2000 | 1.03 | 0.53 | 0.97 | 6.60 | 0.69 | 0.23 | 0.12 |
| 2001 | 6.6 | – | −6.6 | 6.19 | 0.66 | – | 0.11 |
| 2002 | 2.31 | – | 2.31 | 5.49 | 0.66 | – | 0.11 |
| 2003 | 3.46 | 0.16 | 3.46 | 6.33 | 0.70 | 0.33 | 0.03 |
| 2004 | 8.67 | 0.01 | −8.47 | 4.62 | 0.79 | 0.13 | 0.03 |
| 2005 | 9.70 | – | −9.7 | 5.79 | 0.70 | – | 0.14 |
| 2006 | 8.75 | – | −8.75 | 5.23 | 0.73 | – | 0.06 |
| 2007 | 0.85 | 0.66 | −0.57 | 6.55 | 0.70 | 0.35 | 0.11 |
| 2008 | 1.7 | 0.42 | 1.7 | 6.47 | 0.69 | 0.38 | 0.11 |
| 2009 | 2.09 | – | 0.10 | 7.38 | 0.56 | – | 0.19 |
| 2010 | 1.33 | 0.51 | 1.33 | 5.89 | 0.70 | 0.36 | 0.11 |
| 2011 | 3.90 | – | −3.9 | 6.13 | 0.65 | – | 0.12 |
| 2012 | 1.20 | 0.54 | −0.8 | 6.89 | 0.70 | 0.36 | 0.11 |
| 2013 | 1.53 | 0.46 | −1.53 | 5.89 | 0.73 | 0.20 | 0.07 |

not for others because of zero counts of market sentiment states in those years. Using the measure $I_k(f)$, it is clearly seen that the market sentiment states are close to the rational choice of equi-symmetry of $\theta_i$'s for all the years. Turning to the measure $d_\phi(f)$, the market sentiment states are close to the rational choice of equi-symmetry during the entire period of study.

## 5. Conclusions and recommendations

The study of this paper proposes three new measures to analyse absolute market sentiments and to detect the direction of the market sentiments. Evaluating the stock market returns sentiments helps the trader to develop a well-informed and strong trading strategy. In order to evaluate stock returns perception of volatility is critical. Volatility is observed in the daily/weekly/monthly stock returns. The concept of volatility is not easily understood by a trader. To circumvent this, we advocate the analysis of stock returns based on the concept of market sentiments that is very user-friendly and easy to understand by a trader. The focus of this study has been to analyse weekly stock returns applied to market sentiment measures to have a better understanding of the movement in stock returns in terms of market sentiments.

The DCI data analysis shows that the market sentiments are indeed asymmetric and skewed towards the two extreme sentiments of bears and bulls. This is critical for investors who want to minimize risk and maximize profit from equity markets. The investors would therefore know when to trade and when not because they are now equipped with strategies to avoid major losses. Knowing the direction of the market helps one to make prudent resource allocation to avoid dependency on one investment.

The summary findings of the analysis show a continued existence of symmetric market sentiment behaviour witnessed in almost all blocks of years except for a few cases. This proposes that any trader investing in stocks within those years is likely to run a chance of either gaining or losing by the same quantity. Therefore, in case of departure from symmetry a signed measure is applied to market sentiments to determine the direction of the market that could either be bearish or bullish. The analysis shows that for the block of years where the market sentiments witnessed asymmetry towards frowning state of sentiment possibly indicates that the market seems to be under performing.

Our additional findings indicate that investing in the BSE is not yet risky. This declaration of less riskiness has been affirmed by the application of other two measures of equi-symmetry of market sentiments that pronounced symmetric market sentiments behaviour.

In recommendation, the investors in equity markets must be assisted with knowledge about market sentiments states of equity indices, for extreme sentiment states could mean huge losses or gains, resulting in greater uncertainty and thereby impacting on the much needed development of emerging economies like Botswana. We suggest additional listing of domestic companies on BSE as a way of promoting liquidity in the trading of stocks to allow for provision of a friendly environment conducive from those who want to profit from stock trading.

**Author details**
K.K. Moseki[1]
E-mails: mosekikk@mopipi.ub.bw, mosekikk49@gmail.com
ORCID ID: http://orcid.org/0000-0002-3528-8783
K.S. Madhava Rao[1]
E-mail: raom@mopipi.ub.bw
[1] Department of Statistics, University of Botswana, Private Bag 0022, Gaborone, Botswana.

**Citation information**
Cite this article as: Analysing stock market data—Market sentiment approach and its measures, K.K. Moseki & K.S. Madhava Rao, *Cogent Economics & Finance* (2017), 5: 1367147.

**References**
Baker, M. P., & Stein, J. C. (2004). Market liquidity as a sentiment indicator. *Journal of Financial Markets, 7*, 271–299. https://doi.org/10.1016/j.finmar.2003.11.005

Baker, M. P., & Wurgler, J. (2004). *Investor sentiment and the cross-section of stock returns* (NBER Working paper No. 10449). Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w10449

Blake, D. (2005). *Financial market analysis* (2nd ed.). Chichester: Wiley & Sons Ltd.

Bollerslev, A. T. (1987). A conditional heteroscedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics, 69*, 542–547. https://doi.org/10.2307/1925546

Bregman, L. M. (1967). The relaxation method of finding the common points of convex sets and its application to the solution problems in convex programming. *USSR Computational Mathematics and Mathematical Physics, 7*, 200–217.

Brown, G. W. (1999). Volatility, sentiment and noise traders. *Financial Analysts Journal, 55*, 82–90. https://doi.org/10.2469/faj.v55.n2.2263

Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance, 11*, 1–27. https://doi.org/10.1016/j.jempfin.2002.12.001

Cressie, N., & Read, T. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, 46*, 440–464.

Engle, F. R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 987–1008. https://doi.org/10.2307/1912773

Fisher, K. L., & Statman, M. (2000). Investor sentiment and stock returns. *Financial Analysts Journal, 56*, 16–23. https://doi.org/10.2469/faj.v56.n2.2340

Grant, E. L., & Leavenworth, R. S. (1980). *Statistical quality control*. New York, NY: McGraw Hill.

Gujarati, D. N. (2003). *Basic econometrics*. New York, NY: McGraw Hill.

Hirshleifer, D., & Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *The Journal of Finance, 58*, 1009–1032. https://doi.org/10.1111/1540-6261.00556

Hull, C. J. (2003). *Options, futures and other derivatives* (5th ed.). Englewood Cliffs: Prentice Hall.

Hull, J. C. (2012). *Options, future and other derivatives* (8th ed.). Englewood Cliffs: Pearson Education.

Islam, M. A. (2013). Estimating volatility of stock index returns by using symmetric GARCH models. *Middle-East Journal of Scientific Research, 18*, 991–999.

Lee, C. M. C., & Kumar, A. (2006). Retail investor sentiment and return co-movements. *Journal of Finance, 56*, 2451–2486.

Lee, W. Y., Jiang, C. X., & Indro, D. (2002). Stock market volatility, excess returns, and the role of investor sentiment. *Journal of Banking and Finance, 26*, 2277–2299. https://doi.org/10.1016/S0378-4266(01)00202-3

Rao, M. S., & Moseki, K. K. (2009). A probalistic approach to categorise equity index volatility. *Management Dyanamics, 18*, 22–34.

Rao, M. S., & Moseki, K. K. (2011). Analysing volatility in equity indices—A Markov approach for botswana domestic company indices. *South African Journal of Industrial Engineering, 22*, 83–98.

Rao, K. S., & Ramachandran, A. (2014). Exchange rate market sentiment analysis of major global currencies. *Open Journal of Statistics, 4*, 49–69. https://doi.org/10.4236/ojs.2014.41006

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 623–656. https://doi.org/10.1002/bltj.1948.27.issue-4

Sheldon, N. (1994). *Option volatility & pricing: Adavanced trading strategies and techniques*. New York, NY: McGraw Hill Professional.

Shiller, R. J. (1999). Human behavior and the efficiency of the financial system. In J. Taylor & M. Woodford (Eds.), *Handbook of macroeconomics* (pp. 1305–1340). Amsterdam: Elsevier B.V.

Wang, Y. H., Keswani, A., & Taylor, S. (2006). The relationships between sentiment, returns and volatility. *International Journal of Forecasting, 22*, 109–123. https://doi.org/10.1016/j.ijforecast.2005.04.019

Yoshinaga, C. E., & de Castro, Jr., F. F. (2012). The relationship between market sentiment index and stock rates of return: A panel data analysis. *BAR, Rio de Janeiro, 9*, 189–210.

*Cogent Economics & Finance* (ISSN: 2332-2039) is published by Cogent OA, part of Taylor & Francis Group.

**Publishing with Cogent OA ensures:**

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

**Submit your manuscript to a Cogent OA journal at www.CogentOA.com**