Journal on Big Data
DOI:10.32604/jbd.2021.018703
Article

Can Twitter Sentiment Gives the Weather of the Financial Markets?

Imen Hamraoui* **and Adel Boubaker**

International Finance Group Tunisia, Faculty of Management and Economic Sciences of Tunis Tunisia, El Manar University, Tunis Cedex, C. P. 2092, Tunisia

*Corresponding Author: Imen Hamraoui. Email: imen-hamraoui@live.fr

Received: 06 May 2021; Accepted: 17 August 2021

Abstract: Finance 3.0 is still in its infancy. Yet big data represents an unprecedented opportunity for finance. The massive increase in the volume of data generated by individuals every day on the Internet offers researchers the opportunity to approach the question of financial market predictability from a new perspective. In this article, we study the relationship between a well-known Twitter micro-blogging platform and the Tunisian financial market. In particular, we consider, over a 12-month period, Twitter volume and sentiment across the 22 stock companies that make up the Tunindex index. We find a relatively weak Pearson correlation and Granger causality between the corresponding time series over the entire period.

Keywords: Twitter; investor sentiment; tunisian financial market; Twitter volume

1 Introduction

The development of new technologies is leading to a massive increase in the volume, speed, variety and veracity of data available to researchers. The increasing use of the Internet as a source of information has triggered a similar growing online activity. Interaction with technological systems generates massive datasets that document collective behavior in ways previously unimaginable [1]. In this vast repository of Internet activity, academics and practitioners have been able to find the interests, concerns and intentions in order to take advantage of the "Big Data revolution" to bring new perspectives on very diverse issues.

Among the many areas of data research, analysis and modeling, our research is oriented towards a financial systems case study. We believe that investor sentiment as measured by social media is particularly useful for understanding developments in financial markets. On a practical level, the intrinsic complexity of the financial system is mainly due to contagion, caused by collective phenomena such as investor mimicry [2]. Consequently, the possibility of anticipating investor behavior is of paramount interest for economic decision-makers [3] in order to intervene quickly, if necessary.

Over the past decade, several academic researchers have focused on analyzing the relationship between social networks and financial markets. When it comes to social media, Twitter is becoming an increasingly popular microblogging platform used for financial forecasting [4–5].

In this direction, an interesting research by [6] investigates the relationship between the volume of tweets and the financial markets. They tested whether the daily number of tweets predicts the S&P 500 stock market indicators. In a textual analysis approach to Twitter data, the authors find clear relationships between mood indicators and the Dow Jones Industrial Average (DJIA).

In a recent study, Ranco et al. [7] study the relationships between a micro-blogging platform "Twitter" and the financial markets; they find a significant dependence between Twitter sentiment and abnormal returns during volume peaks of Twitter. This is valid not only for expected volume spikes from Twitter (e.g., quarterly announcements), but also for spikes corresponding to less obvious events.

Souza et al. [8] Show that Twitter sentiment for five retail companies has a statistically significant relationship with stock returns and volatility. A recent study by [9] examines the link between investor attention and Bitcoin returns, trade volume and volatility achieved through linear and non-linear Granger causality tests. The result shows that the number of tweets is an important factor in the volume of transactions and the volatility achieved.

2 Motivation

In this article, we study the relationship between stock returns in the Tunisian financial market and the sentiment expressed in financial tweets posted on Twitter. We analyze a set of carefully collected and annotated tweets on the 22 stocks. For each of these companies, we build a time series of sentiment expressed in tweets, with daily resolution. We calculate the Pearson correlation between the price time series and the sentiment time series generated from the tweets. We also perform a Granger causality test to study the predictive power of the Twitter time series. Considering the entire 12-month period, Pearson's correlation values are low and only four companies in our entire sample pass the Granger causality test.

The specificity of our research lies in the introduction of the "Arabic" language [10–11] and the "English" language [12], that is to say that all the tweets posted in Arabic or in English by the investors were analyzed, something that attracts the various stakeholders to publish and popularize their Twitter accounts.

3 Presentation of the Data

3.1 Market Data

The Tunisian financial market can be defined, in terms of market capitalization, as a narrow market, given the small number of companies listed there; this market is dominated or made up of a very large number of companies forming part of the financial sector. We use a representative sample of the Tunisian market composed of 22 companies which are ADWYA, AMEN BANK, AMS, ATB, ASSAD, ATTIJARI BANK, BIAT, BNA, BTE, CARTHAGE CEMENT, CITY CARS, DELICE HOLDING, ELECTROSTAR, LANDOR, MAGASIN GENERAL, POULINA GROUP HOLDING, SAH, SANIMED, SERVICOM, SOPAT, SOTUMAG, TUNISIA LEASING. The choice of the number of companies (sample) is due to the availability and existence of a Twitter account for each share and to the continuity of stock market data. This allows us to trace our study period where these actions seem more or less active on Twitter 01October 2016 to 30 September 2017.

We take as a sample to study all the companies listed on the Tunisian market with a Professional Twitter account. The accounting and financial data were assembled respectively in the "BVMT" databases and the sample includes all listed securities for which data is available.

The first data source takes into account information on stock price returns. For each share, we determine the time series of daily returns,

$$R = \frac{Pt-Pt-1}{Pt-1} \tag{1}$$

R: is the yield of a share.

P_t is the closing price of the share on day t.

These data are accessible to the public and can be downloaded from the BVMT website.

3.2 Twitter Data

The second source of data derives from the Twitter community and includes relevant tweets, as well as their sentiment. The data was collected manually and instantly from the published tweets, where a search query consists of a tag of the action, mentioning hashtag followed by the name of the action (eg #ADWYA). Go et al. [13] used smileys as a proxy for sentiment labels of 1.6 million tweets. Collecting high quality human-labeled tweets is considerably more expensive. Saif et al. [14] give a survey of eight manually annotated datasets having from 500 to 14,000 labeled tweets.

The data of our study covers a 12-month period (01 October 2016 to 30 September 2017), for which there are only 856 tweets for 22 stocks out of a total of 81 listed stocks.

The shares with a Twitter account are: ADWYA- ATTIJARI BANK- ASSAD-DELICE HOLDING- AMEN BANK- ELECTROSTAR- CITY CARS- LANDOR- TUNISIE LEASING- POULINA GROUP HOLDING- MAGASIN GENERAL- SOTUMAG- BIAT- SAH- ATB- SANIMED- SOPAT- BTE- AMS- BNA- CARTHAGE CEMENT- SOTUMAG.

We adopted the same approach as [7], Twitter sentiment is calculated by distinguishing between negative, neutral and positive tweets. This is done from words that indicate pessimism, optimism, or indifference.

- Volume of tweets, TW_d: the total number of tweets in a day.
- Negative tweets, tw_d^-: the number of negative tweets in a day.
- Neutral tweets, tw_d^0: the number of neutral tweets in a day.
- Positive tweets, tw_d^+: the number of positive tweets in a day.
- Sentiment polarity, P_d: the difference between the number of positive and negative tweets as a fraction of non-neutral tweets [15], $\text{Pd} = \frac{tw_d^+ - tw_d^-}{tw_d^+ + tw_d^-}$.

The importance of this work is not to check whether Twitter can move the market, because in the presence of new information and indicating the position of such a stock market policeman, the answer will be clearly positive, the most interesting in our research is to determine whether the dissemination of information via Twitter than via another "more traditional" means of communication can modify the integration of information in the prices of financial assets (speed of integration, volatility, etc.). by keeping the advantages (more transparency, integrity, manipulation, reduction of information asymmetry and market efficiency) and by considering the disadvantages (risk of Twitter account hacking, risk of sending false information, the difficulty of providing "clear" and "precise" information in 140 characters, risk of market overreaction, etc.). This is why we are going to test our null hypothesis (**H0**) Sentiment polarity and tweet volume have no effect on stock market.

4 Methods

For an investigation of the relation between the Twitter sentiment and stock prices, we apply the Pearson correlation and Granger causality tests. We use the Pearson correlation to measure the linear dependence between P_d and R_d. Given two time series, $\mathrm{X_t}$ and $\mathrm{Y_t}$, the Pearson's correlation coefficient is calculated as:

$$\rho(X,Y) = \frac{\mathrm{X_t\,Y_t} - \langle \mathrm{X_t} \rangle \langle \mathrm{Y_t} \rangle}{\sqrt{(\langle \mathrm{X_t^2} \rangle - \langle \mathrm{X_t} \rangle^2) - (\langle \mathrm{Y_t^2} \rangle - \langle \mathrm{Y_t} \rangle^2)}} \tag{2}$$

where $\langle\cdot\rangle$ is the time average value. The correlation $\rho(X,Y)$ quantifies the linear contemporaneous dependence.

We also perform the Granger causality test [16] to check if the Twitter variables help in the prediction of the price returns. The steps of the procedure applied are summarized as follows [17]:

- Determine if the two time series are non-stationary, by the Augmented Dickey-Fuller (ADF) test and Philipp Perron (PP) test.
- Build a Vector Autoregressive (VAR) model and determine its optimal order by considering three measures: AIC, SC, HQ.
- Fit the VAR model with the selected order from the previous step.
- Perform the impulse responses.

- Perform the Granger causality test.

4.1 Preliminary Tests: Investigating the Characteristics of Yield and Sentiment Polarity Series

Before presenting the results of the parameter estimation, as shown in Table 1, we start first with the descriptive statistics, then with the series stationary tests.

Table 1: Descriptive statistics of the securities forming our sample

Title	Medium	Median	Maximum	Minimum	Std.dev	Skewness	Kurtosis	Jarque–Bera	Prob
ADWYA	–0.005406	0.00000	0.046552	–1.000000	0.064888	–14.43179	221.6945	508906.8	0.00000
AMEN BANK	0.010170	0.000000	3.427767	–1.000000	0.225963	13.50832	211.7548	463392.9	0.00000
AMS	–0.059637	0.000000	0.059459	–1.000000	0.237166	–3.678432	14.64138	1983.373	0.00000
ASSAD	–0.025066	0.000000	2.013699	–1.000000	0.227789	0.049525	37.90012	12738.55	0.00000
ATB	–0.069428	0.000000	0.045455	–1.000000	0.253249	–3.372544	12.47342	1414.405	0.00000
ATTIJARI BANK	0.022797	0.000000	5.460648	–0.060829	0.344770	15.72564	248.5336	640844.4	0.00000
BIAT	0.003020	0.000000	1.476217	–1.000000	0.113237	6.040215	139.7734	197169.9	0.00000
BNA	–0.014675	0.000000	0.060217	–1.000000	0.124127	–7.589776	59.48229	35774.48	0.00000
BTE	–0.204677	0.000000	0.626087	–1.000000	0.409332	–1.395615	3.094484	81.57386	0.00000
CARTHAGE CEMENT	–0.007110	0.000000	0.061224	–1.000000	0.086055	–10.09698	110.8147	125832.6	0.00000
CITY CARS	0.087890	0.000000	6.090909	–1.000000	0.502753	6.774770	86329.72	92.83899	0.00000
DELICE HOLDING	–0.050204	0.000000	0.043916	–1.000000	0.222649	–4.031265	17.29265	2816.265	0.00000
ELECTROSTAR	–0.04556	–0.003436	0.076923	–1.000000	0.203310	–4.361397	20.42517	3971.277	0.00000
LANDOR	–0.067325	0.000000	0.996805	–1.000000	0.267774	–2.778902	11.64629	1104.896	0.00000
MAGASIN GENERAL	–0.114106	0.000000	0.044636	–1.000000	0.321033	–2.399356	6.769504	389.4348	0.00000
POULINA GROUP HOLDING	–0.02186	0.000000	0.076056	–1.000000	0.149076	–6.225689	40.21562	16106.25	0.00000
SAH	0.002869	0.000000	0.514706	–0.067857	0.034121	13.49136	203.4370	427777.8	0.00000
SANIMED	–0.094860	0.000000	0.730769	–1.000000	0.305466	–2.496801	7.965014	518.6019	0.00000
SERVICOM	–0.079123	0.000000	0.075000	–1.000000	0.269610	–3.088084	10.61584	1005.527	0.00000
SOPAT	–0.004202	0.000000	1.235955	–1.000000	0.120401	–0.150467	82.78954	66582.57	0.00000
SOTUMAG	–0.02656	0.000000	8.261307	–1.000000	0.576565	11.61509	171.9497	304166.3	0.00000
TUNISIE LEASING	–0.112253	0.000000	0.054231	–1.000000	0.316034	–2.455898	7.050497	302.3031	0.00000

The descriptive analysis of a time series is mainly based on the analysis of the kurtosis coefficients and skewness coefficients. The first informs us about the degree of concentration of the observations around the mean while the second tells us rather the degree of centering of the observations of the series studied, remember that for the normal law, the asymmetry coefficient is zero and the coefficient kurtosis equals 3.

We mainly notice that for the series of sentiment polarity, that the null hypothesis of normality is rejected. We note first of all that the Kurtosis coefficient is very high, that is to say much greater than or different from 3 (theoretical value of the Kurtosis coefficient for the normal law). This phenomenon of excess kurtosis confirms the strongly leptokurtic nature of the series of stock market returns.

Second, the skewness coefficient is different from 0 (theoretical value of the skewness coefficient for the normal law), we notice that the skewness coefficient, in all cases, is positive for the profitability and sentiment polarity series. this indicates that the distribution of each series is spread to the right. We find that the null hypothesis of normality is rejected for the studied series. As a result, the sentiment polarity does not follow the normal law, which is a general characteristic of financial series.

4.2 Stationary Test

The stationary of a series is a necessary condition in any estimation procedure of a model to prove that this is representative of the phenomenon studied.

To understand the stationary or non-stationary character of the profitability series, and of the sentiment indicators, it is necessary to use the two unit root tests: the increased Dickey and Fuller test (1981) and the Philipp Perron test (1988) to avoid them erroneous estimates.

At this stage of our study, it is essential to analyze the stationary of the series of all the variables studied. Indeed, we must test the null hypothesis of the absence of a unit root. The results of ADF and PP are presented in the Table 2.

Table 2: The results of the ADF and PP tests

Title	ADF			PP			Critical Value		
	With Constant and Trend	With Constant	Neither Constant nor Trend	With Constant and Trend	With Constant	Neither Constant nor Trend	With Constant and Trend	With Constant	Neither Constant nor Trend
ADWYA	–15.87601	–15.85772	–15.77888	–15.87681	–15.85847	–15.77888			
AMEN BANK	–54.78992	–54.94225	–55.06662	–54.78992	–55.17612	–55.20671			
AMS	–7.641691	–7.380355	–6.711203	–17.58588	–17.32761	–16.79857			
ASSAD	–19.19348	–19.26694	–3.678328	–18.91845	–18.98892	–18.17755			
ATB	–17.09697	–17.04349	–8.254064	–17.35367	–17.16509	–15.89057			
ATTIJARI BANK	–8.380221	–8.260792	–7.990164	–445.4413	–451.5891	–465.3003			
BIAT	–9.937207	–9.510088	–9.550221	–28.52266	–27.89457	–27.92616			
BNA	–17.96865	–18.02015	–17.90139	–17.96696	–18.02003	–17.90139			
BTE	–13.99501	–14.02101	–6.289079	–21.57253	–21.62153	–17.85545	–3.995189 –3.427902 –3.137310	–3.456408 –2.872904 –2.572900	–2.574245 –1.942099 –1.615852
CARTHAGE CEMENT	–20.49411	–20.58218	–20.58816	–22.33295	–22.44002	–22.05688			
CITY CARS	–26.12262	–26.03979	–	–27.05843	–26.58622	–20.99886			
DELICE HOLDING	–12.69136	–16.60010	–	–17.24898	–16.60010	–			
ELECTROSTAR	–3.307389	–3.285318	–2.859124	–17.52870	–17.54663	–16.84462			
LANDOR	–17.43112	–17.45570	–3.834815	–17.42079	–17.44326	–16.36027			
MAGASIN GENERAL	–18.04195	–18.00552	–	–18.49173	–18.29247	–16.14906			
POULINA GROUP	–5.143716	–4.835795	–4.612595	–16.47403	–16.07873	–15.83306			

HOLDING						
SAH	–15.98834	–15.83392	–15.75261	–15.98733	–15.83399	–15.75277
SANIMED	–17.55075	–2.898188	–10.11158	–17.53716	–17.41755	–16.16431
SERVICOM	–17.49855	–6.043197	–	–17.49855	–17.34084	–
SOPAT	–15.84108	–15.72648	–15.73916	–15.84151	–15.72648	–15.73916
SOTUMAG	–15.73679	–15.75758	–15.75587	–15.80299	–15.82326	–15.81203
TUNISIE LEASING	–6.463988	–5.924866	–	–15.14035	–14.79801	–14.13171

From the stationary table, the ADF and PP tests carried out on all the actions, we notice that the series used is stationary. Let's start with the ADF test (with constant and trend), this variable presents a value of ADF which is lower than the critical values displayed directly by EVIEWS, at the three thresholds 1%, 5% and 10%. The same observation appears for the PP test, but the trend is not significant, because the t-statistic is less than the critical value which is around (1.96), with a non-zero probability different from 5%. For this reason, we are going to move on to the second model which is the ADF model (with constant), the results seem to say that the sentiment polarity variable is stationary in level, because this variable has a value of ADF which is lower than the values. Critics, Same observation appears for the PP test, with the significance of all the variables, hence the rejection of the null hypothesis H0 "there is a unit root, the process is not therefore stationary" of the PP test. We can notice as well as the probabilities of accepting H0 for all the series in the two stationary tests: ADF and PP is equal to zero, we can conclude that all the series are stationary in level.

4.3 VAR Model Estimation

The VAR model is an econometric model used to capture the interdependencies between the time series and their evolutions, all the variables in the VAR model are treated symmetrically, for the estimation of our VAR model, we start first with the determining the delay order.

4.3.1 Determination of the Delay Number

Based on the three information criteria, namely the modified SCHWARTZ criterion (SC), the AKAIKE criterion and the Hannan-Quinn criterion, which are judged to be efficient, as well as the log likelihood in determining the delay order of vector autoregressive models (VAR), these criteria showed us an optimal delay order equal to 1 in Table 3.

Table 3: The criteria for choosing the optimal delay order of the VAR model

		1	2	3	4
Adwya	AIC	–2.292536*	–2.261380	–2.233186	–2.202283
	SC	–2.207288*	–2.119300	–2.034273	–1.946538
	HQ	–2.258214*	–2.204178	–2.153102	–2.099318
AMEN BANK	AIC	–2.101096*	–2.073325	–2.057984	–2.026948
	SC	–2.015847*	–1.931245	–1.859072	–1.771203
	HQ	–2.066774*	–2.016122	–1.977901	–1.923983
AMS	AIC	–0.175479*	–0.183300	–0.168131	–0.142132
	SC	–0.090230*	–0.041219	0.030782	0.113613
	HQ	–0.141157*	–0.126097	–0.088047	–0.039167
ASSAD	AIC	0.330884*	0.353502	0.337441	0.353560
	SC	0.416132*	0.495583	0.536353	0.609304
	HQ	0.365205*	0.410705	0.417525	0.456524

ATB	AIC SC HQ	–0.661981* –0.576732* –0.627659*	–0.637679 –0.495599 –0.580476	–0.615318 –0.416405 –0.535234	–0.609263 –0.353518 –0.506298
ATTIJARI BANK	AIC SC HQ	–5.870962 –5.785714* –5.836641*	–5.876216* –5.734135 –5.819013	–5.863040 –5.664127 –5.782956	–5.833022 –5.577277 –5.730057
BIAT	AIC SC HQ	–1.557878* –1.472630* –1.523557*	–1.527499 –1.385419 –1.470297	–1.502143 –1.303231 –1.422060	–1.500395 –1.244651 –1.397431
BNA	AIC SC HQ	–0.721739 –0.636491* –0.687418*	–0.724516* –0.582435 –0.667313	–0.717493 –0.518580 –0.637409	–0.689997 –0.434252 –0.587032
BTE	AIC SC HQ	0.070353* 0.155602* 0.104675*	0.080408 0.222489 0.137611	0.109428 0.308341 0.189512	0.140172 0.395917 0.243137
CARTHAGE CEMENT	AIC SC HQ	–2.642032 –2.556784* –2.607710	–2.610153 –2.468073 –2.552951	–2.709822 –2.510909 –2.629738*	–2.716430* –2.460686 –2.613466
CITY CARS	AIC SC HQ	–0.323597* –0.238349* –0.289275*	–0.295732 –0.153651 –0.238529	–0.273002 –0.074089 –0.192918	–0.268307 –0.012562 –0.165342
DELICE HOLDING	AIC SC HQ	–0.056612* 0.028636 –0.022291*	–0.052308 0.089772 0.004895	–0.055939 0.142973 0.024145	–0.028502 0.227243* 0.074463
ELECTROSTAR	AIC SC HQ	–0.140057 –0.054808* –0.105735*	–0.110070 0.032011 –0.052867	–0.143205* 0.055708 –0.063121	–0.116771 0.138974 –0.013806
LANDOR	AIC SC HQ	0.034186* 0.119434* 0.068508*	0.055412 0.197492 0.112615	0.078985 0.277897 0.159069	0.094958 0.350703 0.197923
MAGASIN GENERAL	AIC SC HQ	0.469808* 0.555057* 0.504130*	0.494808 0.636889 0.552011	0.512291 0.711203 0.592374	0.540387 0.796132 0.643352
POULINA GROUP HOLDING	AIC SC HQ	–1.561766 –1.476517* –1.527444*	–1.581995* –1.439914 –1.524792	–1.553776 –1.354863 –1.473692	–1.527278 –1.271533 –1.424313
SAH	AIC SC HQ	–6.079023* –5.993774* –6.044701*	–6.056469 –5.914388 –5.999266	–6.043888 –5.844976 –5.963805	–6.014732 –5.758987 –5.911767

SANIMED	AIC SC HQ	–0.407090 –0.321842 –0.372769	–0.409566* –0.267486 –0.352364	–0.382859 –0.183946 –0.302775	–0.392097 –0.136352 –0.289132
SERVICOM	AIC SC HQ	–0.599158 –0.513909 –0.564836	–0.570343 –0.428263 –0.513141	–0.561894 –0.362982 –0.481811	–0.534585 –0.278841 –0.431621
SOPAT	AIC SC HQ	–1.115281* –1.030032* –1.080959*	–1.084452 –0.942372 –1.027250	–1.089859 –0.890946 –1.009775	–1.084454 –0.828709 –0.981489
SOTUMAG	AIC SC HQ	1.634140* 1.719389* 1.668462*	1.658560 1.800641 1.715763	1.687116 1.886028 1.767199	1.708597 1.964342 1.811562
TUNISIE LEASING	AIC SC HQ	0.880860 0.989367* 0.924873 *	0.873417 1.054262 0.946773	0.815401* 1.068584 0.918099	0.845731 1.171252 0.977772

4.3.2. The Estimation Results of the VAR Model

The Table 4 presents the estimation of the parameters by the VAR model, and shows the existing relationship between the different variables of the model, it shows the results of the entire sample of the bi-varied VAR model of the performance of Tunisian companies listed in stock market and the polarity of investor sentiment. For each coefficient, we have the estimated value, the standard error and the statistical *t* value.

Table 4: VAR model estimate per share

		Rp(–1)	SP(–1)	C
Adwya	**Rp**	–0.006957 (0.06363) [–0.10934]	–0.000860 (0.01446) [–0.05950]	–0.005463 (0.00414) [–1.31866]
	SP	–0.030358 (0.27837) [–0.10906]	0.110215 (0.06324) [1.74279]	0.001912 (0.01813) [0.10547]
AMEN BANK	**Rp**	0.005203 (0.01814) [0.28681]	0.002752 (0.01343) [0.20495]	–0.003684 (0.00415) [–0.88728]
	SP	–0.004525 (0.08650) [–0.05231]	0.019869 (0.06402) [0.31036]	0.044440 (0.01980) [2.24451]
AMS	**Rp**	–0.075825 (0.06145) [–1.23398]	0.096983 (0.06415) [1.51176]	–0.062615 (0.01509) [–4.15058]
	SP	0.022152 (0.06074) [0.36471]	–0.086007 (0.06341) [–1.35632]	0.023044 (0.01491) [1.54538]
ASSAD	**Rp**	–0.007827 (0.05245) [–0.14925]	0.003161 (0.03343) [0.09456]	–0.033760 (0.01256) [–2.68880]
	SP	0.060366 (0.09976) [0.60509]	0.032400 (0.06360) [0.50945]	0.106023 (0.02388) [4.43915]
ATB	**Rp**	–0.075440 (0.06317)	0.046113 (0.09459)	–0.073977 (0.01668)

		[–1.19422]	[0.48751]	[–4.43582]
	SP	–0.109530	0.086903	0.008180
		(0.04178)	(0.06257)	(0.01103)
		[–2.62130]	[1.38898]	[0.74150]
ATTIJARI BANK	**Rp**	0.183457	0.003386	0.000541
		(0.06206)	(0.00224)	(0.00067)
		[2.95590]	[1.51250]	[0.80914]
	SP	2.082775	0.122780	0.034984
		(1.76932)	(0.06381)	(0.01906)
		[1.17716]	[1.92412]	[1.83534]
BIAT	**Rp**	–0.002032	0.003939	–0.003483
		(0.03626)	(0.00974)	(0.00435)
		[–0.05603]	[0.40444]	[–0.80097]
	SP	–0.128601	0.186959	0.123551
		(0.23262)	(0.06247)	(0.02789)
		[–0.55285]	[2.99278]	[4.42953]
BNA	**Rp**	–0.005859	–0.018034	–0.009627
		(0.05674)	(0.01927)	(0.00723)
		[–0.10328]	[–0.93607]	[–1.33121]
	SP	–0.144643	0.105107	0.071847
		(0.18793)	(0.06382)	(0.02395)
		[–0.76967]	[1.64704]	[2.99934]
BTE	**Rp**	–0.261834	–0.255393	–0.265892
		(0.06063)	(0.16099)	(0.02773)
		[–4.31823]	[–1.58639]	[–9.58982]
	SP	–0.014558	–0.176285	–0.021812
		(0.02363)	(0.06273)	(0.01080)
		[–0.61616]	[–2.81003]	[–2.01884]
CARTHAGE CEMENT	**Rp**	–0.011645	0.004004	–0.003875
		(0.04978)	(0.01760)	(0.00428)
		[–0.23391]	[0.22742]	[–0.90491]
	SP	–0.431654	0.102995	0.018027
		(0.17209)	(0.06085)	(0.01480)
		[–2.50834]	[1.69250]	[1.21791]
CITY CARS	**Rp**	–0.034874	0.069530	–0.116797
		(0.03985)	(0.12999)	(0.02042)
		[–0.87524]	[0.53487]	[–5.71909]
	SP	–0.009554	–0.011771	0.015345
		(0.01950)	(0.06363)	(0.01000)
		[–0.48989]	[–0.18501]	[1.53515]
DELICE HOLDING	**Rp**	–0.053548	0.030658	–0.054171
		(0.06352)	(0.05700)	(0.01463)
		[–0.84295]	[0.53785]	[–3.70261]
	SP	0.006577	0.082868	0.030903
		(0.07069)	(0.06343)	(0.01628)
		[0.09304]	[1.30639]	[1.89807]
ELECTROSTAR	**Rp**	–0.114161	0.160640	–0.049549
		(0.06040)	(0.04240)	(0.01248)
		[–1.89013]	[3.78847]	[–3.96905]
	SP	–0.231826	0.260552	–0.001722
		(0.08778)	(0.06162)	(0.01814)
		[–2.64102]	[4.22804]	[–0.09492]
LANDOR	**Rp**	–0.069834	0.077675	–0.080756
		(0.06128)	(0.07116)	(0.01740)

		[–1.13965]	[1.09157]	[–4.64036]
	SP	–0.061032	0.088445	0.048137
		(0.05444)	(0.06322)	(0.01546)
		[–1.12112]	[1.39906]	[3.11349]
MAGASIN GENERAL	**Rp**	–0.132563	–0.021104	–0.129242
		(0.06311)	(0.09301)	(0.02164)
		[–2.10046]	[–0.22689]	[–5.97305]
	SP	–0.040877	–0.095970	0.025620
		(0.04463)	(0.06577)	(0.01530)
		[–0.91595]	[–1.45911]	[1.67447]
POULINA GROUP HOLDING	**Rp**	–0.029398	–0.065411	–0.020501
		(0.06375)	(0.04813)	(0.00965)
		[–0.46117]	[–1.35896]	[–2.12441]
	SP	–0.140425	0.068271	0.022734
		(0.07978)	(0.06024)	(0.01208)
		[–1.76011]	[1.13328]	[1.88225]
SAH	**Rp**	–0.005312	0.002741	0.002773
		(0.06361)	(0.00834)	(0.00221)
		[–0.08350]	[0.32855]	[1.25224]
	SP	–0.163980	0.084033	0.044434
		(0.48337)	(0.06339)	(0.01683)
		[–0.33924]	[1.32567]	[2.64038]
SANIMED	**Rp**	–0.099099	0.073630	–0.106026
		(0.06334)	(0.13646)	(0.02037)
		[–1.56453]	[0.53957]	[–5.20590]
	SP	0.021435	–0.018856	0.025060
		(0.03223)	(0.06944)	(0.01036)
		[0.66506]	[–0.27156]	[2.41814]
SERVICOM	**Rp**	–0.096823	0.059281	–0.088440
		(0.06336)	(0.10115)	(0.01796)
		[–1.52809]	[0.58609]	[–4.92526]
	SP	0.022483	–0.088753	0.025739
		(0.03972)	(0.06341)	(0.01126)
		[0.56598]	[–1.39962]	[2.28634]
SOPAT	**Rp**	0.000333	0.002529	–0.004372
		(0.06461)	(0.02743)	(0.00794)
		[0.00515]	[0.09218]	[–0.55079]
	SP	0.420361	–0.048143	0.078637
		(0.14977)	(0.06360)	(0.01840)
		[2.80673]	[–0.75703]	[4.27398]
SOTUMAG	**Rp**	–0.001090	–0.055295	–0.024168
		(0.06363)	(0.16396)	(0.03741)
		[–0.01713]	[–0.33724]	[–0.64601]
	SP	–0.019355	0.120011	0.038204
		(0.02449)	(0.06310)	(0.01440)
		[–0.79044]	[1.90204]	[2.65368]
TUNISIE LEASING	**Rp**	–0.128223	–0.052843	–0.093223
		(0.07326)	(0.07881)	(0.02883)
		[–1.75023]	[–0.67050]	[–3.23332]
	SP	0.088320	0.134584	0.047836
		(0.06913)	(0.07436)	(0.02720)
		[1.27767]	[1.80980]	[1.75838]

According to Table 4, we see a dependence between the polarity of sentiment and the profitability of some securities given the existence of the important coefficients, same remark for the polarity of sentiment which depends on the delayed value of a period of return, These values are greater than the critical value which is of the order of 1.96 at a level of 5%, one can interpret this by the existence of the coefficients relatively close to zero which explain the effect of the sentiment polarity on the return at a delay level equal to 1.

Indeed, the sentiment polarity of the securities (ATB- CARTHAGE CEMENT- ELECTROSTAR and POULINA GROUP HOLDING) have an absolute t-statistic greater than or close to 1.96 which are respectively of the order of (2.62; 2.508; 2.64; 1.76) these securities depend negatively on the lagged performance of a period. Since their coefficients are negative (-0.109; -0.431; -0.2318; -0.14). However, the SOPAT security depends positively with a t-statistic (2.806) on the delayed return of a security period with a coefficient (0.42).

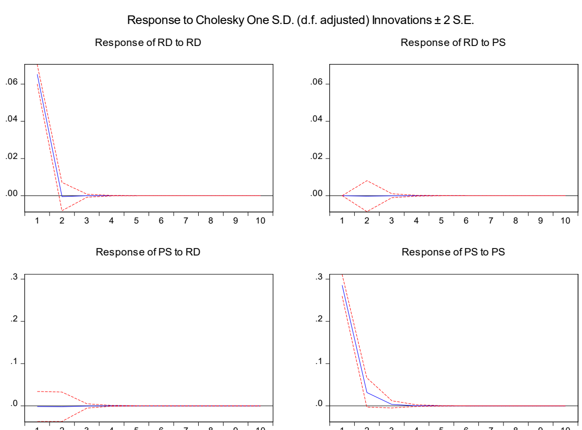
Likewise, we find that the return depends positively on the polarity of delayed sentiment of a period for a single stock in our sample. This title (ELECTROSSTAR) records a t-statistic which is worth (3.78) which is greater than 1.96 with a positive coefficient equal (0.16).

This brings us back to saying that there is a relationship between (R_{it}) and (SP). Also, we observe the positive effect of the variable (R_{it}) at the level of the delay (1) on itself at the instant t is presenting positive t-statistic values.

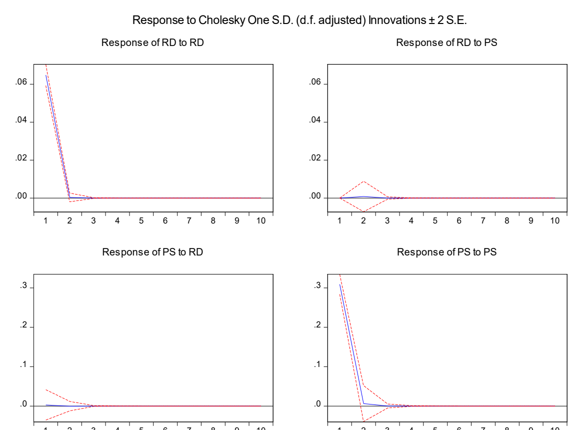
We can conclude that, generally there are the stocks more sensitive to investor sentiment than others. This result is consistent with the study by [18] who showed that investor sentiment simulates securities in cross-section. They show that some companies are more refined in investor sentiment than others.

4.4 Impulse Response Function

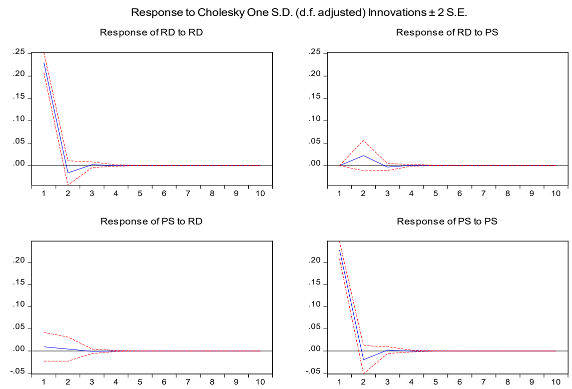
This function marks the effect of a shock on the current and future values of endogenous variables. For our VAR model, this describes a relationship between the performance of Tunisian companies, and the sentiment polarity. We seek to identify the impact of stock return at date t, according to the dynamics of the variable (SP) at periods subsequent to t and vice versa, assuming the evolutions of these two variables for $t \leq T$ known and data, according to the graphs of these coefficients and the associated values, we can see that we find the general appearance of the response function to an innovation. Estimates of VAR coefficients cannot capture the full impact of an observation of the endogenous variable. To do this, the impulse response functions use all the estimates of the VAR coefficients, to plot the full impact of a residual shock. Fig. 1 contains the possible graph of impulse response function using the estimate of VAR.



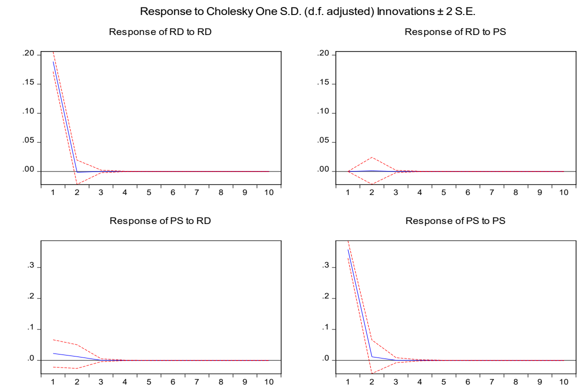
ADWYA



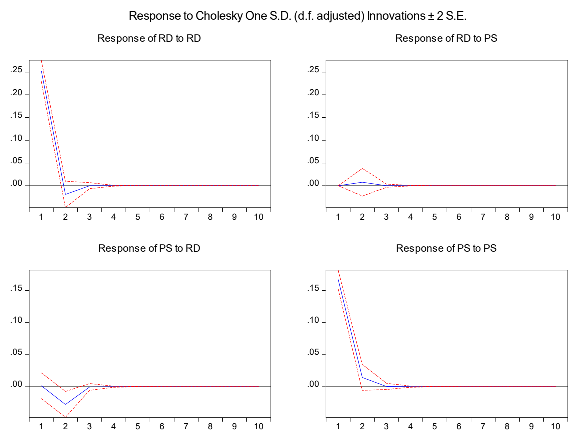
AMEN BANK



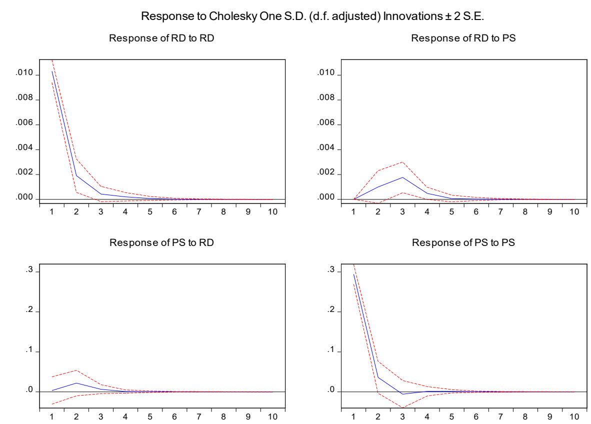
AMS



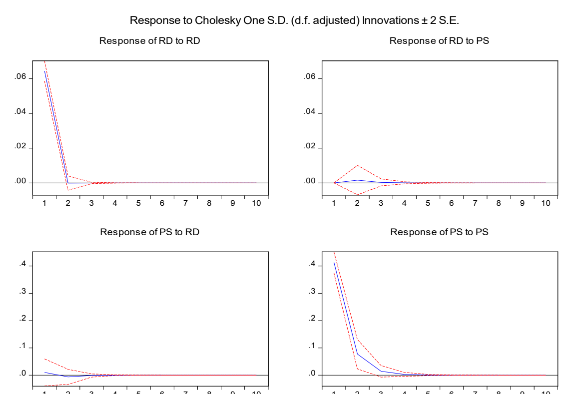
ASSAD



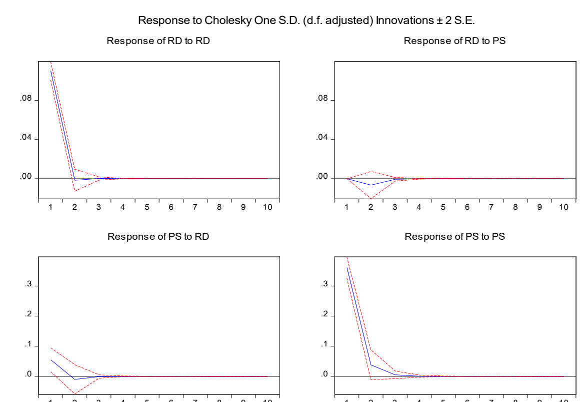
ATB



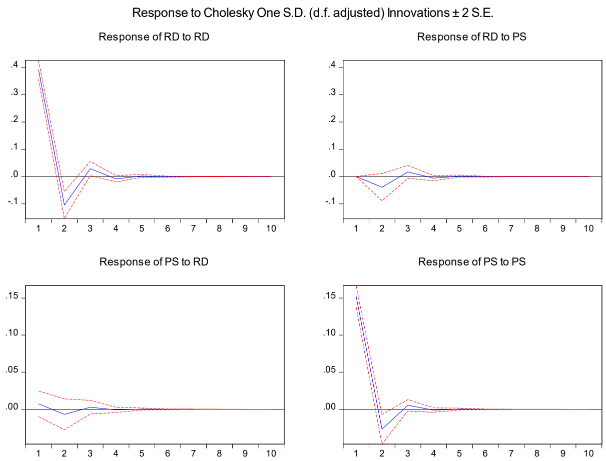
ATTIJARI BANK



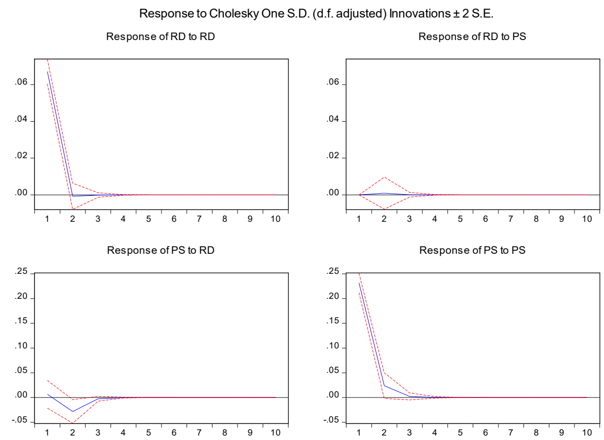
BIAT



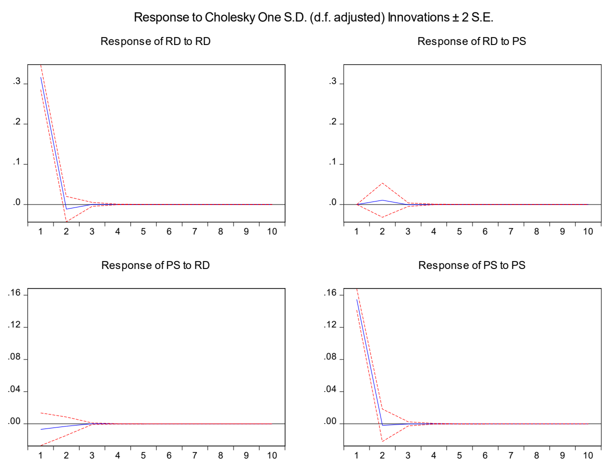
BNA



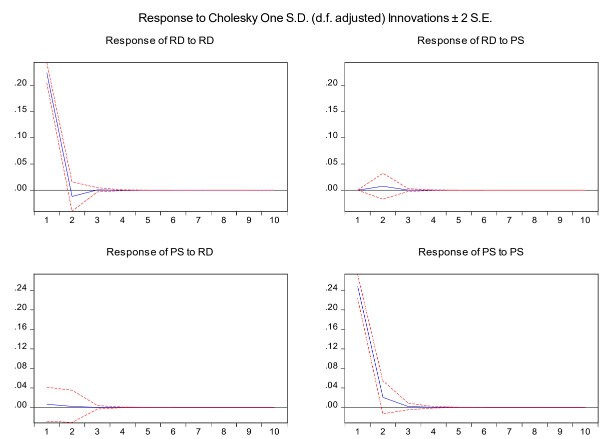
BTE



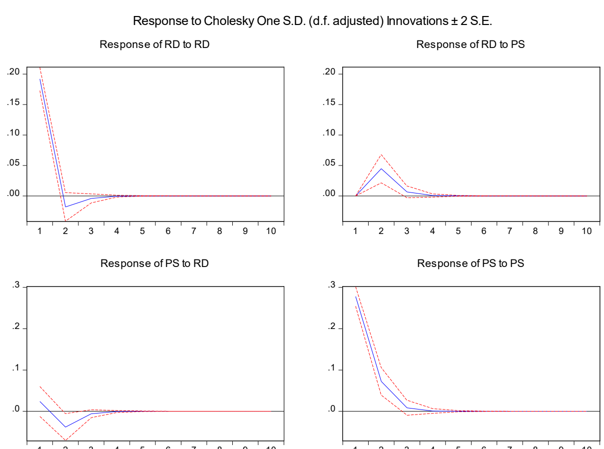
CARTHAGE CEMENT



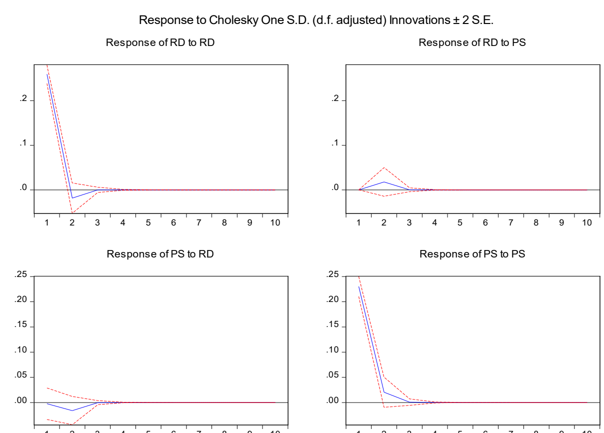
CITY CARS



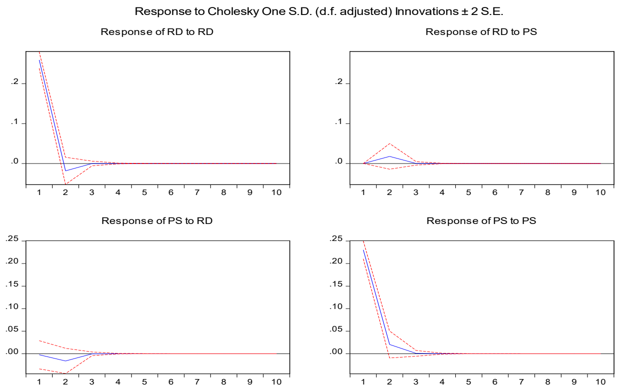
DELICE HOLDING



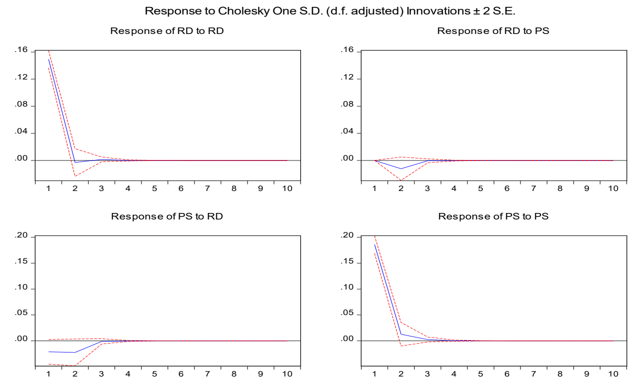
ELECTROSTAR



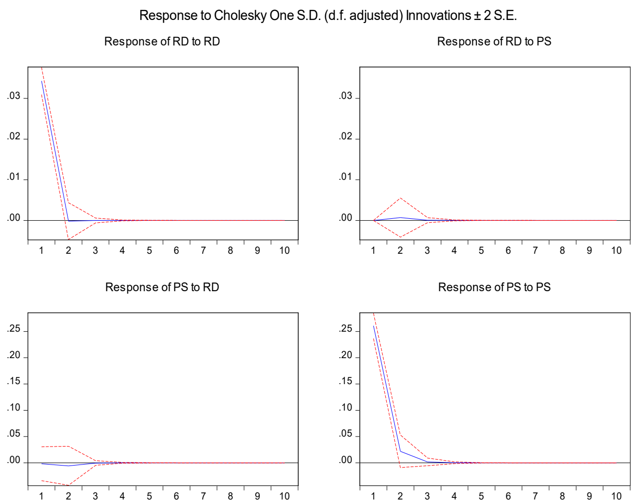
LANDOR



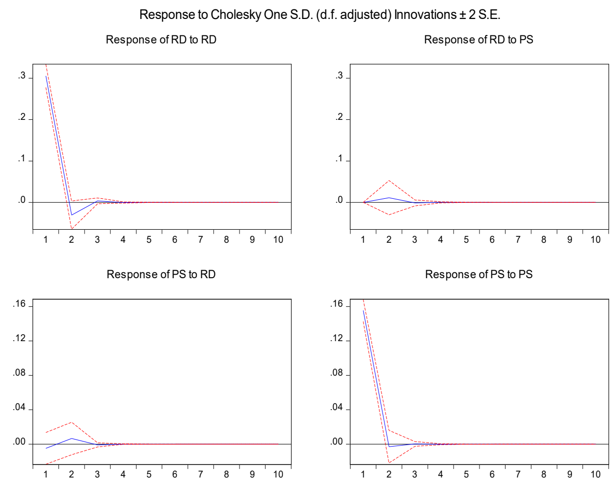
MAGASIN GENERAL



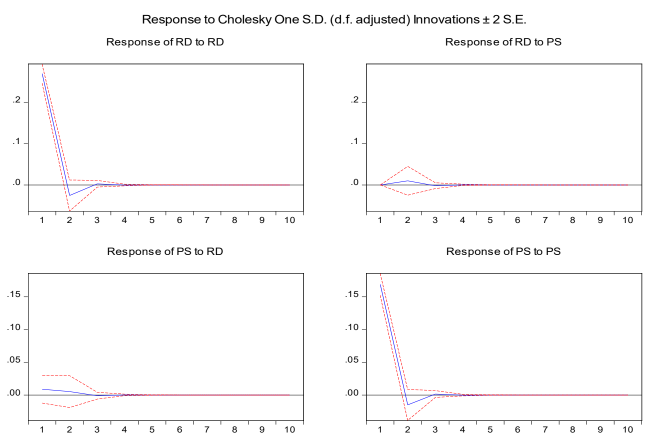
POULINA GROUP HOLDING



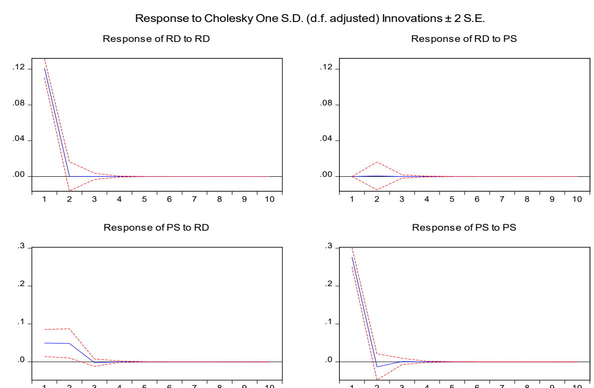
SAH



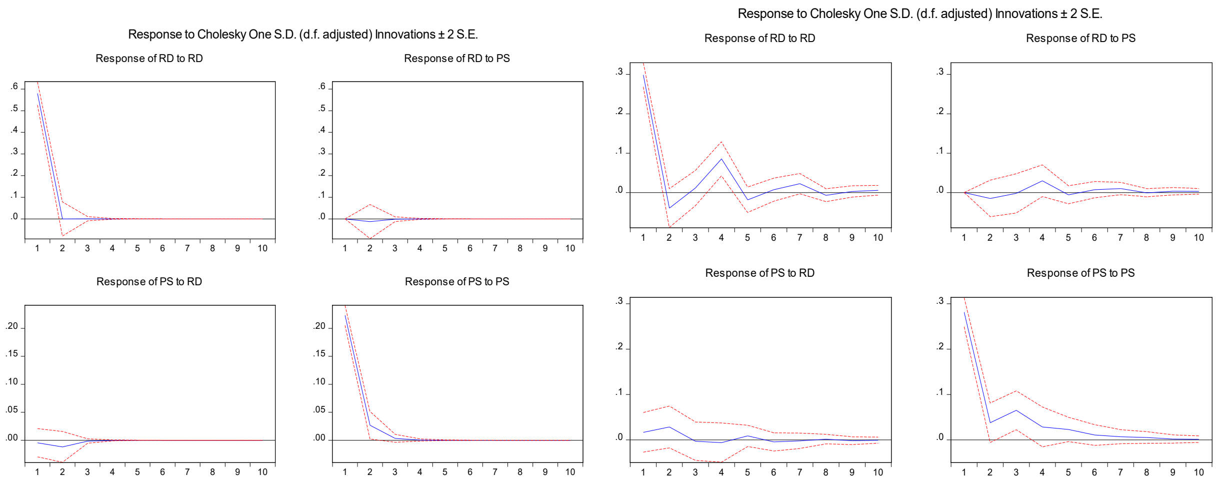
SANIMED



SERVICOM



SOPAT



SOTUMAG TUNISIE LEASING

Figure 1: Impulse responses of stock return and sentiment polarity for different stocks

Table 5: Granger causality tests

	The TESTS							
Title	RD does not Granger Cause SP		SP does not Granger Cause RD		WT does not Granger Cause RD		RD does not Granger Cause WT	
	F-Statistic	Prob.	F-Statistic	Prob.	F-Statistic	Prob.	F-Statistic	Prob.
Adwya	0.05316	0.9482	0.02054	0.9797	0.13117	0.8771	0.16762	0.8458
AMEN BANK	0.45275	0.6364	0.11136	0.8947	0.48570	0.6159	10.2783	**5.E-05**
AMS	0.33289	0.7172	1.21400	0.2988	0.47134	0.6247	1.25844	0.2859
ASSAD	0.44497	0.6414	0.28484	0.7524	0.34981	0.7052	0.63673	0.5299
ATB	3.53456	**0.0307**	0.20574	0.8142	0.46742	0.6272	0.75393	0.4716
ATTIJARI BANK	0.70821	0.4935	4.02248	**0.0191**	0.21595	0.8059	0.05016	0.9511
BIAT	1.88672	0.1538	0.10020	0.9047	0.07639	0.9265	3.46779	**0.0327**
BNA	0.37414	0.6883	0.80845	0.4467	0.00374	0.9963	0.44586	0.6408
BTE	0.94336	0.3907	1.52459	0.2198	0.16758	0.8458	2.35210	**0.0973**
CARTHAGE CEMENT	0.03056	0.9699	0.04805	0.9531	0.02337	0.9769	0.06480	0.9373
CITY CARS	0.36403	0.6952	0.33327	0.7169	5.81008	**0.0034**	0.12594	0.8817
DELICE HOLDING	0.11221	0.8939	0.28951	0.7489	1.03223	0.3578	0.98757	0.3740
ELECTROSTAR	3.43028	**0.0339**	6.85664	**0.0013**	3.19333	**0.0428**	2.87136	**0.0585**
LANDOR	1.08565	0.3393	0.58838	0.5560	0.28583	0.7516	0.19570	0.8224
MAGASIN GENERAL	0.47815	0.6205	0.20727	0.8129	0.91394	0.4023	0.45459	0.6352
POULINA GROUP HOLDING	1.57653	0.2088	0.32875	0.7201	0.30844	0.7349	0.29732	0.7431

SAH	1.77554	0.1716	1.38738	0.2517	0.22868	0.7958	9.21042	**0.0001**
SANIMED	0.65502	0.5203	0.48034	0.6192	0.64331	0.5264	0.72715	0.4843
SERVICOM	0.40859	0.6650	0.18391	0.8321	0.42003	0.6575	0.74078	0.4778
SOPAT	3.88551	**0.0218**	0.05820	0.9435	0.13501	0.8738	4.65298	**0.0104**
SOTUMAG	0.33103	0.7185	0.41033	0.6639	1.03754	0.3559	0.14496	0.8651
TUNISIE LEASING	0.99932	0.3703	0.16666	0.8466	0.65495	0.5208	1.03077	0.3589

The first two graphs (1) and (2) (R to R and R to SP) note the yield response to a shock of a sentiment polarity, respectively on the variable itself. The vertical axis in these two graphs measures the percentage increase in R. For graphs (3) and (4) (SP to R and SP to SP), they plot the response of sentiment polarity to a shock d standard deviation, respectively on the yield and on the variable itself. The graphs as well as the tables of the impulse responses of the innovations, can inform us about the purpose of our study, because of the representation of the responses to the shocks of (R) and (SP) for the later periods.

On the one hand, the daily profitability of Tunisian companies shows a zero response during the first subsequent period, then a slight increase, following a shock in the sentiment polarity during the second and the third subsequent period, then a return to point d. equilibrium during the fourth posterior period by a process of correction, also, Fig. 1 indicates a slight decrease in the polarity of feeling at a shock of (R), this response gradually decreases and returns to the point of equilibrium by a process of correction during the second later period, this leads us to reveal a node of our hypothesis. This result is interpreted as obvious that the sentiment polarity has no effect on the returns of Tunisian companies and vice-versa.

On the one hand, the effect of a shock of the sentiment polarity on the profitability of a security presents a zero response for all subsequent periods for the following securities (ADWYA- AMEN BANK- ASSAD- CARTHAGE CEMENT-MAGASIN GENERAL and SOPAT) however this shock shows a slight increase, during the second and third subsequent period, then a return to equilibrium for the fourth period for securities (AMS- ATB- BIAT- CITY CARS- DELICE HOLDING- ELECTROSTAR-LANDOR - SAH).

4.5 Granger Causality Tests

Granger's causality tests [16] are performed to verify the direction of causality between the returns of different securities and investor sentiment. This test is used to examine whether there is a positive relationship between sentiment polarity and Tunisian stock market. To do this, we must use the causality test of Granger (1969,1980) because it examines a double causality between two variables, that is to say a set of processes (X_{jt}) with j = (1… q), if there is a link of causality between (X_{kt}) and (X_{jt}), this means that we can better predict (X_{jt}) using the past values of (X_{kt}).

In Granger's sense, a series of profitability "causes" the series of sentiment polarity or volume of tweets if past knowledge of profitability improves prediction of sentiment polarity or the number of tweets and/or vice versa. Thus, causality tests make it possible to highlight the direction of the causal relationship between the profitability of Tunisian listed companies and the sentiment polarity and the volume of tweets. Can taking SP and Wt into account improve profitability forecasting and/or reciprocally?

Test1: H0: Sentiment Polarity (SP) does not predict profitability in Granger's sense

Test2: H0: The volume of tweets (WT) does not predict in Granger's sense the returns of securities.

From the Table 5, we tried to study the Granger causal links by company between the sentiment polarity and stock market on the one hand, and the volume of tweets and stock market on the other hand.

It can be seen that the return on the securities does not cause in the sense of Granger the sentiment is accepted since its probability is relatively high (0.05) except for four companies which have probabilities of less than 5% which are ATB- ATTIJARI BANK and SOPAT and which is respectively (0.0307-

0.0191- 0.0218) these companies prove that there is a causal relation in only one direction. The ELECTROSTAR company illustrates the existence of a double causality in the sense of Granger between the return and the sentiment polarity seen that its probability in both directions is (0.0339- 0.0013).

The number of tweets for a company seems essential for more than 30% of the companies in our sample. Indeed, these actions have probabilities of less than 5% which are AMEN BANK- BIAT- BTE- CITY CARS- SAH and SOPAT and which is respectively (5.E-05- 0.0327- 0.0973- 0.0001- 0.0104) these companies prove that there is a causal relationship in one direction. The ELECTROSTAR company illustrates the existence of a double causality in the sense of granger between the return and the sentiment polarity seen that its probability in both directions is (0.0428- 0.0428).

5 Discussion

In this work we present dependence between stock price returns and Twitter sentiment in tweets about the companies. As a series of other papers have already shown, there is a signal worth investigating which connects social media and market behavior [7].

The conclusions that can be drawn are as follows:

• The sentiment polarity variable is not very useful for predicting stock performance because only four Tunisian companies in our entire sample pass the Granger test.

• The number of tweets variable is important because paying attention to Twitter is useful in predicting price volatility.

From the estimation of the VAR model, the Granger causality test and the impulse response functions, it can be concluded that there is a relatively weak correlation between sentiment polarity and stock profitability. Sentiment polarity is a non-significant variable since there is no relationship between the two variables. However, the variable of the number of tweets is partially significant from which it can be concluded that accounting for the volume of tweets is important in predicting price volatility. This result is in favor of behavioral finance.

According to our results, sentiment has predictive power for returns. This proves that emotions play an essential role in explaining the evolution of stock prices. They are often referred to as a state of mind. Traditional financial theory suggests that a project is profitable only when investors make their choices based solely on their end goals. Unlike this discipline, behavioral or so-called emotional finance assumes that the more an investor is aware of his emotions, the more he makes better decisions. This means that emotions support an individual's ability to make rational decisions.

Our results corroborate those obtained by [7]. These authors examine, over a 15-month period, the volume and opinion of Twitter for 30 stock companies in the Dow Jones Industrial Average (DJIA). Their results prove a relatively weak Pearson correlation and Granger causality between the corresponding time series over the entire period. Indeed, the conclusions that can be drawn are twofold; on the one hand, the polarity variable is superfluous to predict the price return because only three companies pass the Granger test. On the other hand, the number of tweets for one company causes performance volatility for a third of companies. This indicates that Twitter is essential for forecasting price volatility.

Another similar result developed by [6], who studied over a period of 4 months with daily data (from 16 February 2012 to 10 May 2012) with a sample of four sectors (finance, energy, health and materials) whether the daily number of tweets that mention S & P500 stocks can predict stock market indicators of S&P 500 stocks in the stock market. Their results show that for each day, the number of tweets correlates with stock market information. Additionally, it seems that Twitter is useful for predicting the stock market.

Baker et al. [18] examine the impact of investor sentiment on the cross-sectional distribution of equity returns. Their starting hypothesis is that investor confidence will have significant effects on particular securities that are difficult to arbitrate. Following this prediction, their result seems to say that when confidence indicators at the start of the period are low, subsequent returns are relatively high for stocks belonging to the following category (small, young stocks, stocks with high volatility, unprofitable stocks, non-dividend paying stocks, extreme growth stocks, and distressed stocks). However, when sentiment is high, these classes of shares produce relatively low subsequent returns. The authors conclude their study with the negative effect of investor sentiment on the cross section of stock prices. To this end, they suggest that at the asset valuation level, the expected return model should incorporate investor sentiment.

Our results also appear similar in part to those reported by [19]; these authors analyze around 250,000 stock-related tweets on a daily basis. They find an association between tweet sentiment and stock market returns, this contradicts our result, and however, they show a significant relationship between the volume of tweets and the performance of a stock.

Taken together, this research asserts that the publication volume contains information that can be used by and in the financial markets. They turn out to be an additional source of information that can adequately anticipate dung movements.

By way of conclusion, our study reveals a negative or even zero relationship between the sentiment polarity and the stock market; on the other hand, we have proven that the number of tweet has a significant effect on the return on a share. We therefore conclude that twitter is a financial forecasting barometer. We tried to present a few theorists who found the same result as our study as well as other researchers who found contradictory results

6 Conclusion

In our empirical work, we studied the effect of information published via the social network "Twitter" on the profitability of shares. The most interesting thing in our research is to determine whether the disclosure of information via Twitter than via another "more traditional" means of communication can modify the integration of information in the prices of financial assets (speed of integration, volatility…). To do this, it seems interesting to test the impact of sentiment polarity as well as the volume of tweet on the performance of a security.

Two sources of daily data were used, the first source relating to the market for which the daily performance of the Tunindex was used, the second source relating to Twitter data.

Our sample is made up of 22 listed companies over a period extending from 01 October 2016 to 30 September 2017. The choice of our sample and the study period is supported by the presence of an active professional Twitter account of these different companies throughout the study period.

We have already applied the same sentiment classification methodology proposed by [7] and, inspired by the study of [19–20].

From the Granger causality test, we concluded that the Sentiment Polarity variable is not very useful for predicting stock returns because only four Tunisian companies in our entire sample pass the Granger test. Nonetheless, the number of tweets variable (the volume of tweets) is important because paying attention to twitter is useful in predicting price volatility.

Acknowledgement: The authors would like to thank the International Finance Group of Tunisia research laboratory as well as all those who contributed to carry out this study.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] G. King, "Ensuring the data-rich future of the social sciences," *Science*, vol. 331, pp. 719–721, 2011.

[2] J. P. Bouchaud, "The (unfortunate) complexity of the economy," *Physics World*, pp. 28–32, 2009.

[3] J. P. Bouchaud, "Economics needs a scientific revolution," *Nature*, vol. 455, no. 48, pp. 291–292, 2008.

[4] A. Java, X. Song, T. Finin and B. Tseng, “Why we twitter: understanding microblogging usage and communities,” in *Proc. of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, pp. 56–65, 2007.

[5] M. Graham, S. A. Hale and D. Gaffney, “Where in the world are you? Geolocation and language identification in Twitter,” *The Professional Geographer*, vol. 66, no. 4, pp. 568–578, 2014.

[6] Y. Mao, W. Wei, B. Wang and B. Liu, “Correlating S&P 500 stocks with Twitter data,” in *1st ACM Int. Workshop on Hot Topics on Interdisciplinary Social Networks Researchp*, pp. 69–72, 2012.

[7] G. Ranco, A. Darko, C. Guido, G. Miha and M. Igor, “The effects of Twitter sentiment on stock price returns,” *PLoS One*, pp. 1–21, 2015.

[8] T. T. P. Souza, O. Kolchyna, P. C. Treleaven and T. Aste, “Twitter sentiment analysis applied to finance: A case study in the retail industry,” *Business, Compute Science, Economic*, 2015.

[9] D. Shen, A. Urquhart and P. Wang “Does Twitter predict Bitcoin?” *Economics Letters*, vol. 174, pp. 118–122, 2019.

[10] M. Pontiki, G. Dimitris, P. Haris, A. Ion, M. R. Suresh *et al.,* “SemEval-2016 task 5: Aspect based sentiment analysis,” in *Proc. of the 10th Int. Workshop on Semantic Evaluation*, pp. 19–30, 2016.

[11] S. Mohammad, S. Kiritchenko, P. Sob-hani, X. D. Zhu and C. Cherry. “SemEval-2016 task 6: Detecting stance in tweets,” in *Proc. of the 10th Int. Workshop on Semantic Evaluation*, pp. 31–41, 2016.

[12] S. Rosenthal, N. Farra and P. Nakov, “SemEval-2017 task 4: Sentiment analysis in Twitter,” in *Proc. of the 11th Int. Workshop on Semantic Evaluations*, pp. 502–518, 2017.

[13] A. Go, R. Bhayani and L. Huang, “Twitter sentiment classification using distant supervision,” *Cs224n Project Report*, 2009.

[14] H. Saif, M. Fernandez, Y. He and H. Alani, “Evaluation datasets for Twitter sentiment analysis,” in *Proc. of the 1st Workshop on Emotion and Sentiment in Social and Expressive Media*, 2013.

[15] W. Zhang and S. Skiena, “Trading strategies to exploit blog and news sentiment,” in *Proc. of the Fourth Int. Conf. on Weblogs and Social Media*, 2010.

[16] C. W. J. Granger, “Investigating causal relations by econometric models and cross-spectral methods,” *Econometrica: Journal of the Econometric Society*, pp. 424–438, 1969.

[17] M. Piškorec, N. Antulov-Fantulin, P. K. Novak, I. Mozetič, M. Grčar *et al.,* “Cohesiveness in financial news and its relation to market volatility,” *Scientific Reports*, vol. 4, no. 1, pp. 1–8, 2014.

[18] M. Baker and W. Jeffrey, “Investor sentiment and the Cross-Section of stock returns,” *Journal of Finance*, vol. 61, no. 4, pp. 1–36, 2006.

[19] T. O. Sprenger, P. G. Sandner, A. Tumasjan and I. M. Welpe, “News or noise? using Twitter to identify and understand company-specific news flow,” *Journal of Business Finance & Accounting*, vol. 41, no. 7–8, pp. 791–830, 2014.

[20] T. O. Sprenger, A. Tumasjan, P. G. Sandner and I. M. Welpe, “Tweets and trades: The information content of stock microblogs,” *European Financial Management*, vol. 20, no. 5, pp. 926–957, 2014.