

The effect of COVID-19 on the Egyptian exchange using principal component analysis

Heba M. Ezzat

Faculty of Economics and Political Science, Cairo University, Giza, Egypt

Abstract

Purpose – Since the beginning of 2020, economies faced many changes as a result of coronavirus disease 2019 (COVID-19) pandemic. The effect of COVID-19 on the Egyptian Exchange (EGX) is investigated in this research.

Design/methodology/approach – To explore the impact of COVID-19, three periods were considered: (1) 17 months before the spread of COVID-19 and the start of the lockdown, (2) 17 months after the spread of COVID-19 and the during the lockdown and (3) 34 months comprehending the whole period (before and during COVID-19). Due to the large number of variables that could be considered, dimensionality reduction method, such as the principal component analysis (PCA) is followed. This method helps in determining the most individual stocks contributing to the main EGX index (EGX 30). The PCA, also, addresses the multicollinearity between the variables under investigation. Additionally, a principal component regression (PCR) model is developed to predict the future behavior of the EGX 30.

Findings – The results demonstrate that the first three principal components (PCs) could be considered to explain 89%, 85%, and 88% of data variability at (1) before COVID-19, (2) during COVID-19 and (3) the whole period, respectively. Furthermore, sectors of food and beverage, basic resources and real estate have not been affected by the COVID-19. The resulted Principal Component Regression (PCR) model performs very well. This could be concluded by comparing the observed values of EGX 30 with the predicted ones (R -squared estimated as 0.99).

Originality/value – To the best of our knowledge, no research has been conducted to investigate the effect of the COVID-19 on the EGX following an unsupervised machine learning method.

Keywords Dimensionality reduction, EGX 30, Principal component analysis, Multiple imputation

Paper type Research paper

1. Introduction

Since the end of Year 2019, economies faced many challenges due to the imposed lockdown and public fear. Financial markets play significant role in countries' economies (Mishkin, 2010). However, financial markets are very fragile to sudden changes (Cont, 2001). Henceforth, inspecting the effect of coronavirus diseases 2019 (COVID-19) on financial markets is of significant importance.

Many studies were run to investigate the effect of the pandemic on the stock market of different countries [c.f. (Amin *et al.*, 2021; Hong *et al.*, 2021; Uddin *et al.*, 2021; Yousfi *et al.*, 2021; He *et al.*, 2020; Sharif *et al.*, 2020)]. Awad (2020) and Elayed and Abdelrhim (2020) investigated the impact of number of COVID-19 cases on the EGX. Authors above studied the contribution of individual stocks to EGX 30 (which includes the major 30 individual stocks in terms of activity and liquidity traded in the EGX) to identify stable sectors in the EGX.



To our knowledge, no study has been conducted to investigate the effect of COVID-19 on the EGX following unsupervised machine learning approach. In this paper, results are drawn from the stock trading behavior instead of linking number of COVID-19 cases to the EGX. By this, any inefficiency or inaccuracy in announced number of COVID-19 cases are avoided. This paper aims to help detecting the most affected individual stocks by COVID-19. Results of this research will aid investors to decide on which stocks to trade during and pre the period of COVID-19. Also, this research would help the government to identify the negatively affected sectors to support them. Finally, researchers can study the spillover effect on the EGX returns.

The PCA is a dimensionality reduction technique, which main goal is to express original variables with a much smaller number of PCs that explain most of the data variability. A data set with n variables can be converted into a new lower dimension set of p (where, $p < n$) principal components (PCs) following the PCA. Each PC is a linear combination of the original n variables. PCA addresses the two main challenges in financial time series: (1) multicollinearity and (2) the huge number of dimensions (30 stocks).

Recently, much research has been developed different models that could work efficiently for financial market prediction (Hargreaves, 2019; Zhong and Enke, 2019; Cavalcante *et al.*, 2016). As stock markets could be affected by a vast number of factors, it seems to be a wise idea to rely on the most important variables to predict the future behavior. Many researchers developed prediction models following dimensionality reduction techniques [c.f. (Cao and Wang, 2020; Ghorbani and Chong, 2020; Zhang, 2018; Waqar *et al.*, 2017)]. In this research, the future behavior of the EGX is predicted following the PCR model, which is a regression analysis relying on the PCA algorithm. The period included in the study composes a period before the pandemic and a period since the start of the spread of COVID-19.

The main results reveal that the first three PCs could be considered to explain most of the data variability at (1) before COVID-19, (2) during COVID-19 and (3) the whole period. Moreover, sectors of food and beverage, basic resources and real estate have not been affected by the COVID-19. The developed PCR model performs very well. This could be decided by comparing the observed values of EGX 30 against the predicted ones (R -squared estimated as 0.99).

The rest of the paper is divided to five sections. Section 2 reviews the literature of studying the effect of COVID-19 on financial markets. Section 3 describes the data and their preprocessing. Section 4 illustrates the research methodology and the PCA and PCR algorithms. Main results and analyses are displayed in Section 5. Finally, Section 6 concludes the paper.

2. Literature review

In this section, we are going to review the literature of previous studies that investigated the effect of COVID-19 on financial markets. Baker *et al.* (2020) studied the effect of COVID-19 on the US stock market. They concluded that the crash caused by COVID-19 was very dramatic compared to previous crashes caused by infectious diseases.

Additionally, Mazur *et al.* (2021) investigated the US stock market performance during the COVID-19 pandemic. The main findings show that the natural gas, food, health care, and software stocks earned high returns. On the other hand, petroleum, real estate, entertainment and hospitality sectors fell drastically. Liu *et al.* (2021) studied the effect of COVID-19 on the stock market in China. They found that the pandemic increased stock crash risk.

Izzeldin *et al.* (2021) investigated the impact of COVID-19 on the stock markets in the G7 countries. The authors found that the health care and consumer services were the most severely affected sectors. The technology sector was marginally hit, as enforced lockdown pushed people to exploit web-based entertainment and other distraction sources.

Contessi and Pace (2021) studied the spread of COVID-19 and stock market collapses in 18 major countries. The main results reveal that the instability was transmitted from the stock market in China to all other markets, especially the European ones.

Amin *et al.* (2021) studied the effect of the spread of COVID-19 on financial markets in three regions: Central America, North America and South America. The results show that COVID-19 had negative impact on the stock markets. However, there was an insignificant correlation between COVID-19 and the stock market in South America.

Hong *et al.* (2021) investigated the association between COVID-19 and the instability of both stock return predictability and price volatility in the US. The results display a single break in return predictability and price volatility. Also, the timing of the break is synchronized with the COVID-19 outbreak. Uddin *et al.* (2021) studied the effect of COVID-19 pandemic on stock market volatility and economic strength. The effect was evaluated by a set of selected country-level economic measures and factors. The results show that governments could minimize financial volatility by setting different economic policy responses.

Yousfi *et al.* (2021) performed a comparative assessment of the effect of the first and second waves of COVID-19 for the US stock market and its uncertainty. The paper presents dynamic conditional correlation and asymmetric impacts of waves on the association between the US and the Chinese stock markets before and during COVID-19. Also, the correlation between COVID-19 and the US returns and uncertainty during the pandemic were examined. The results support the existence of spillover effect between the two stock markets. Also, a persistent link between the US returns, uncertainty, and COVID-19 was observed. The results demonstrate that the pandemic caused harmful effects on financial markets in general and on the US economy in specifically.

He *et al.* (2020) studied the spillover effect of COVID-19 on stock markets, such as China, Italy, South Korea, France, Spain, Germany, Japan and the US. The results illustrate that COVID-19 has a negative but short-term effect on stock markets of affected countries. Additionally, the effect of COVID-19 on stock markets has bidirectional spillover association between Asian countries and European and American countries.

Given the foregoing discussion and argument, we speculated the following hypothesis:

H1. Some sectors in the EGX are affected by COVID-19 more than others.

To examine *H1*, we divided the period under investigation to three windows: before COVID-19, during COVID-19 and the whole period. The correlation between individual stocks contributing to each PC and the respective PCs was examined.

3. Data description and preprocessing

This research investigates the effect of COVID-19 on the EGX. To achieve research goal, we studied the EGX 30 price index and its constituents in the period from August 1, 2018 to April 30, 2021 (670 observations). This period is divided to two main windows: (1) before COVID-19 (from August 1, 2018 to January 1, 2020) and (2) since spread of COVID-19 (from January 1, 2020 to April 30, 2021). There were 20 individual stocks registered in EGX 30 for the two-time windows. Description of individual stocks' names, symbols and sectors is provided in Table 1.

As illustrated in Table 1, some data were missing. Deleting their corresponding records may generate a totally biased data set. Accordingly, the Missing Completely At Random (MCAR) Test was run. The *p*-value for the Hawkins test of normality and homoscedasticity is 2.35e-26. This implies that either the test of multivariate normality or homoscedasticity (or both) is rejected. Provided that normality can be assumed, the hypothesis of MCAR is rejected at 0.05 significance level. Multiple Imputation by Chained Equations (MICE) approach that uses Classification And Regression Trees (CART) was followed to replace the missing values (Burgette and Reiter, 2010). It does not set parametric assumptions or data transformations to

Company	Stock	Number of missing data (Percentage)	Sector
Arab Cotton Ginning	ACGC	–	Personal and Household Products
Oriental Weavers	ORWE	–	
Eastern Company	EAST	–	
Citadel Capital – Common Shares	CCAP	–	Financial Services excluding Banks
Egyptian Financial Group-Hermes Holding Company	HRHO	–	
Egyptian Kuwaiti Holding	EKHO	–	Banks
Pioneers Holding	PIOH	1 (0.60%)	
Commercial International Bank (Egypt)	COMI	1 (0.15%)	
Credit Agricole Egypt	CIEB	21 (3.13%)	Chemicals
Egyptian Chemical Industries (Kima)	EGCH	–	
Juhayna Food Industries	JUFO	38 (5.70%)	Food and Beverage
Egyptian for Tourism Resorts	EGTS	6 (0.90%)	Travel and Leisure
Egyptian Iron and Steel	IRON	2 (0.30%)	Basic Resources
Ezz Steel	ESRS	–	Industrial Goods and Services and Automobiles Real Estate
ELSWEDY ELECTRIC	SWDY	–	
Heliopolis Housing	HELI	1 (0.15%)	
Medinet Nasr Housing	MNHD	–	
Palm Hills Development Company	PHDC	–	
Six of October Development and Investment (SODIC)	OCDI	1 (0.15%)	
T M G Holding	TMGH	–	

Table 1.
Description of the 20
individual stocks listed
in EGX 30 during the
investigated period

fit nonlinear relations and complex distributions. The algorithm was run for 1,000 times, and the median of these runs was considered to replace the missing values. Figure 1 displays the result of the multiple imputation.

Heteroskedasticity is one of the important features of financial time series that would influence the results of the PCA. Thereafter, stock returns are used for the PCA instead of stock closing prices. Stock returns, r_{tj} , are calculated as

$$r_{tj} = \left(\frac{p_{tj} - p_{(t-1)j}}{p_{(t-1)j}} \right) \quad (1)$$

where $p_{tj} = \log(P_{tj})$ and P_{tj} is the closing price at time t , where $t = 1, 2, \dots, T$, for stock j , where $j = 1, 2, \dots, J$.

4. Research methodology

Many studies show that Stock market data are highly correlated (Sharma and Banerjee, 2015; Solnik *et al.*, 1996). One important feature of the PCA is the ability to address multicollinearity and high dimensionality. The PCA algorithm is run after data preprocessing. Data suitability for the PCA are tested by Kaiser–Meyer–Olkin (KMO) Statistic. The lowest (highest) value of the KMO is 0.81 (0.97) for IRON (EAST) stock. The KMO-criterion for the whole data set is 0.92 disclosing the capability to perform the PCA.

The major aim of the PCA technique is to project the data into new directions. Thus, the main purpose of the PCA is finding lower dimensions (less than the number of original variables included in the model) that can explain most of the data variability. The data set

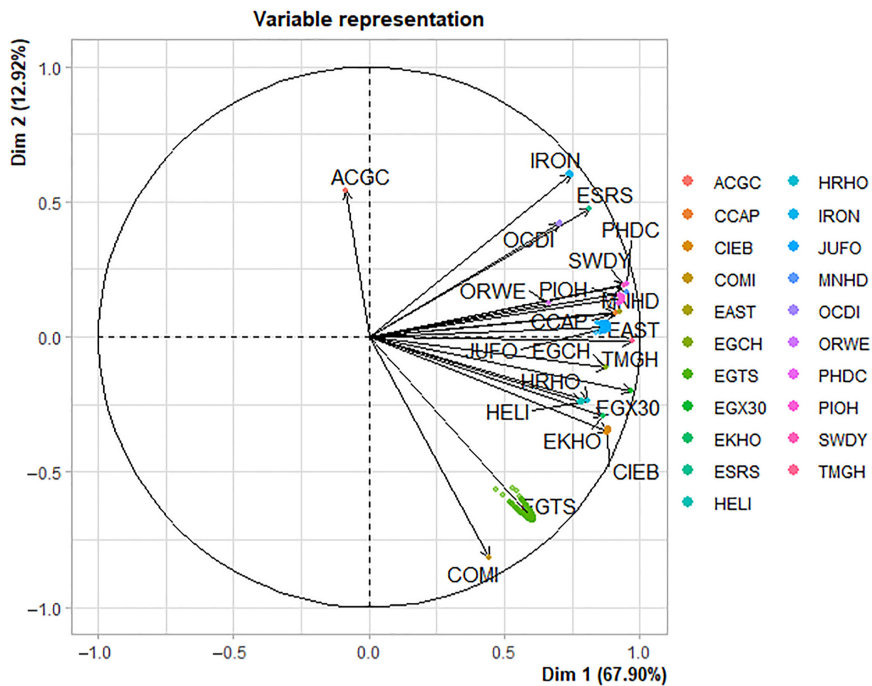


Figure 1.
Multiple imputation for
individual stock prices
with missing data

under study includes 20 variables. PCA will generate, optimistically, lower number of PCs (less than 20) than can explain most of the data variability (at least 70%). The new dimensions or PCs are orthogonal, so they are uncorrelated.

To find the PCs, we need first to standardize the data, such as

$$z_{ij} = \frac{r_{ij} - \bar{r}_j}{s_j} \tag{2}$$

where $\bar{r}_j$ is the mean and s_j is the standard deviation for stock j , where $j = 1, 2, \dots, J$.

Then, the correlation matrix is calculated for standardized values. The PCs are the eigenvectors of the correlation matrix. These eigenvectors are the directions that explain most data variance. Eigenvalues are the coefficients associated with the eigenvectors. Eigenvalues provide the amount of variance that could be captured by each PC.

5. Results and analyses

The main results of the PCA are represented in this section. Tables 2–4 report the eigenvalue, percentage of explained variance and cumulative percentage of variance for each PC for the time windows: (1) before COVID-19, (2) during COVID-19 and (3) the whole period, respectively. To decide on the number of PCs that should be considered, three rules in the literature could be followed (Jolliffe, 2002): (1) consider the first PCs that explain at least 70 percent of cumulative variance, (2) in the scree plot determine at which PC the percentage of explained variance would not be improved (the elbow) and/or (3) choosing PCs with eigenvalues that are greater than one, this rule is known as Kaiser’s rule (Kaiser, 1960). Tables 2–4 report that the first three PCs could explain about 89%, 85% and 88%,

Table 2.Summary of main
results of the PCA
before COVID-19

PCs	Eigenvalue	% of var.	Cumulative % of var.	PCs	Eigenvalue	% of var.	Cumulative % of var.
Dim.1	3.95	42.89	42.89	Dim.11	0.03	0.33	99.16
Dim.2	3.22	34.91	77.8	Dim.12	0.02	0.18	99.35
Dim.3	1.03	11.21	89.01	Dim.13	0.01	0.15	99.5
Dim.4	0.31	3.33	92.34	Dim.14	0.01	0.14	99.64
Dim.5	0.25	2.69	95.03	Dim.15	0.01	0.11	99.75
Dim.6	0.11	1.22	96.26	Dim.16	0.01	0.09	99.84
Dim.7	0.09	0.98	97.24	Dim.17	0.01	0.06	99.89
Dim.8	0.06	0.64	97.88	Dim.18	0.01	0.06	99.95
Dim.9	0.05	0.55	98.43	Dim.19	0	0.03	99.98
Dim.10	0.04	0.4	98.83	Dim.20	0	0.02	100

Table 3.Summary of main
results of the PCA
after COVID-19

PCs	Eigenvalue	% of var.	Cumulative % of var.	PCs	Eigenvalue	% of var.	Cumulative % of var.
Dim.1	4.78	50.31	50.31	Dim.11	0.05	0.48	98.26
Dim.2	2.09	22.04	72.36	Dim.12	0.04	0.41	98.67
Dim.3	1.21	12.7	85.06	Dim.13	0.03	0.34	99.01
Dim.4	0.47	4.96	90.02	Dim.14	0.02	0.24	99.25
Dim.5	0.2	2.16	92.18	Dim.15	0.02	0.18	99.43
Dim.6	0.16	1.72	93.9	Dim.16	0.02	0.16	99.59
Dim.7	0.13	1.34	95.24	Dim.17	0.01	0.15	99.74
Dim.8	0.1	1.07	96.31	Dim.18	0.01	0.1	99.84
Dim.9	0.08	0.87	97.18	Dim.19	0.01	0.09	99.93
Dim.10	0.06	0.6	97.78	Dim.20	0.01	0.07	100

Table 4.Summary of main
results of the PCA for
the whole period

PCs	Eigenvalue	% of var.	Cumulative % of var.	PCs	Eigenvalue	% of var.	Cumulative % of var.
Dim.1	13.3	66.61	66.61	Dim.11	0.09	0.46	98.61
Dim.2	2.67	13.35	79.97	Dim.12	0.06	0.31	98.92
Dim.3	1.58	7.89	87.86	Dim.13	0.04	0.22	99.14
Dim.4	0.6	3.02	90.88	Dim.14	0.04	0.2	99.34
Dim.5	0.54	2.69	93.57	Dim.15	0.04	0.18	99.51
Dim.6	0.29	1.44	95.01	Dim.16	0.03	0.15	99.66
Dim.7	0.23	1.17	96.18	Dim.17	0.03	0.13	99.79
Dim.8	0.16	0.8	96.98	Dim.18	0.02	0.09	99.88
Dim.9	0.12	0.61	97.59	Dim.19	0.02	0.08	99.96
Dim.10	0.11	0.57	98.15	Dim.20	0.01	0.04	100

respectively, of data variability. This cumulative percentage of variance is considered very pleasing.

Figure 2 illustrate the scree plot for the three times windows: before COVID-19, during COVID-19 and the whole period, respectively. The figure shows that after the fourth PC the added explanation for the variance is very low. Additionally, Tables 2–4 report that eigenvalues attached to the first three PCs are greater than one. Accordingly, the first three PCs can explain variance more than the variance that could be explained by single original variables. So, the three rules support reflecting the first three PCs in the analysis.

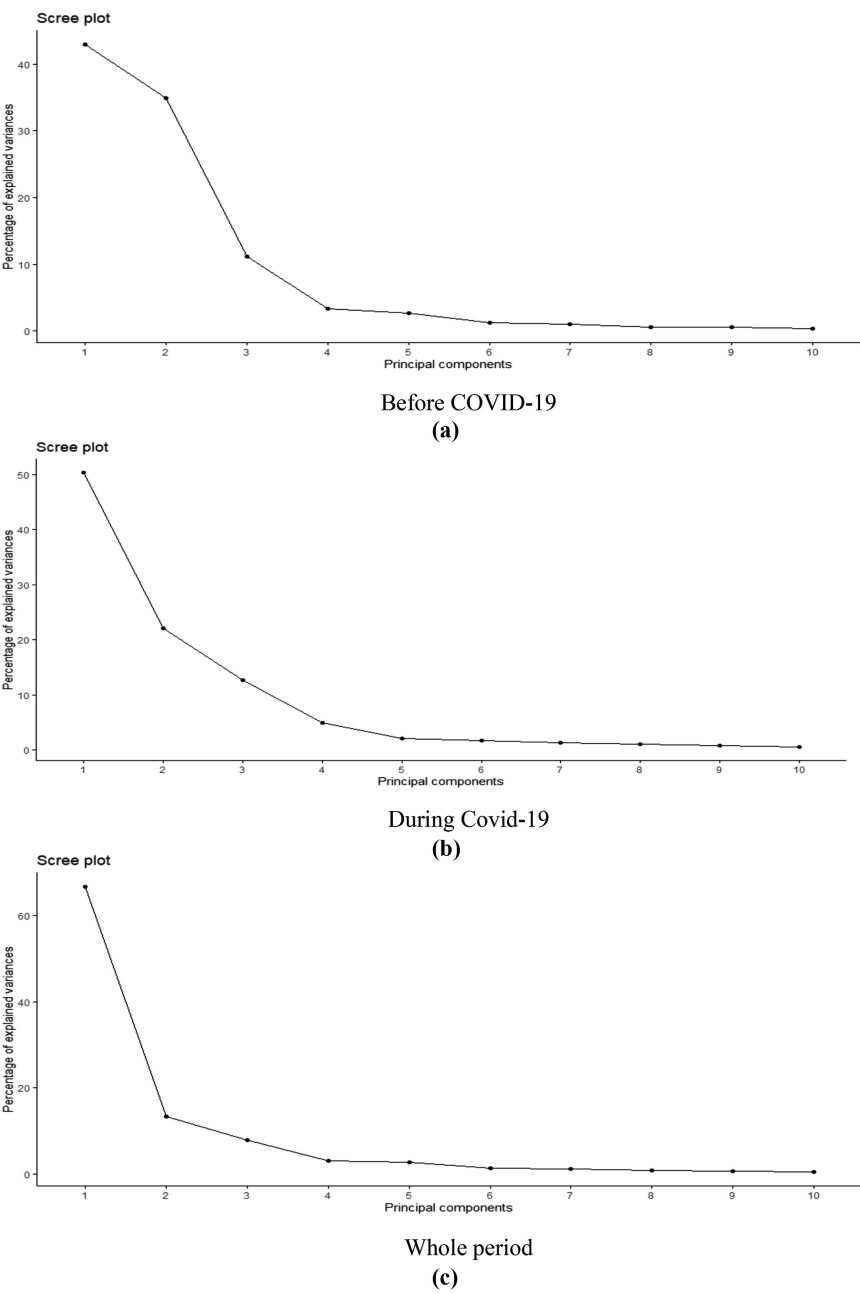
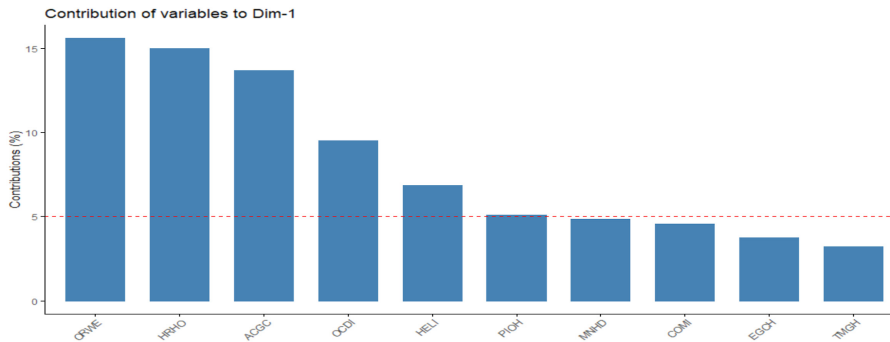
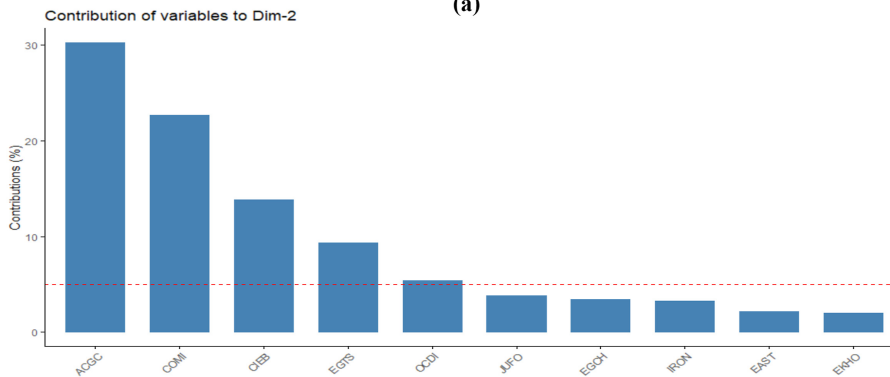


Figure 2. Scree plot depicting the percentage of variance explained by each PC in before COVID-19, during COVID-19 and the whole time period, respectively

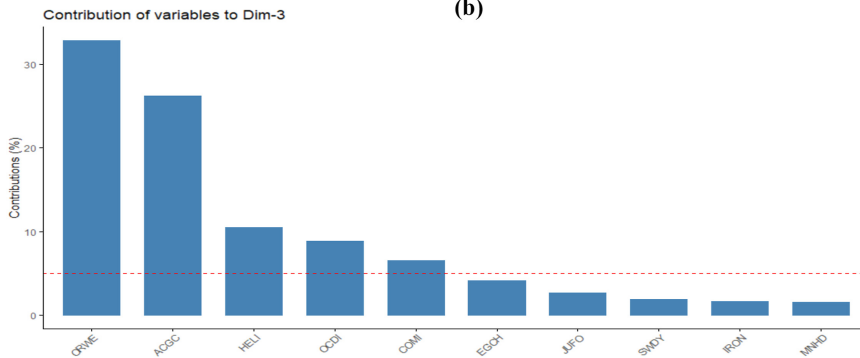
Each panel in Figures 3–5 display the highest ten individual stocks contributing to the first three PCs for each time window: before COVID-19, during COVID-19 and the whole period, respectively. For the before COVID-19 time window, ORWE, HRHO, ACGC, OC DI and HELI are the highest stocks contributing to the first PC; ACGC, COMI, CIEB and EGTS are



Before COVID-19
(a)



During Covid-19
(b)



Whole time-period
(c)

Figure 3.
The most ten
contributing stocks to
the first three PCs
before COVID-19,
respectively

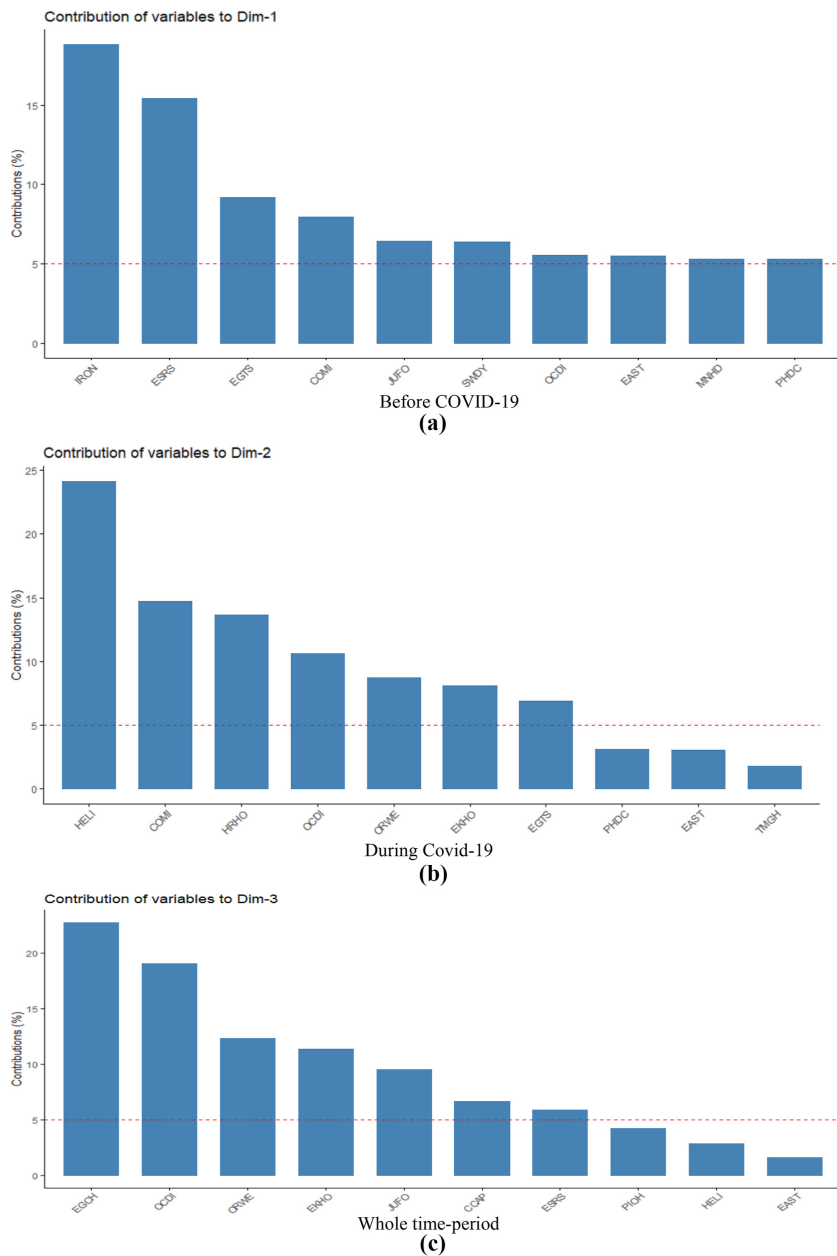
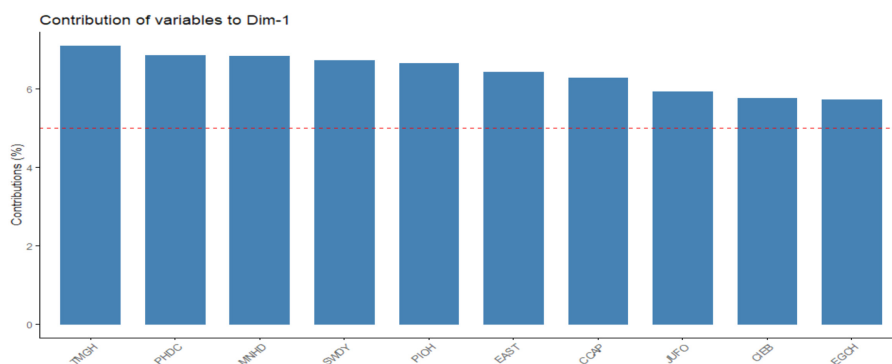
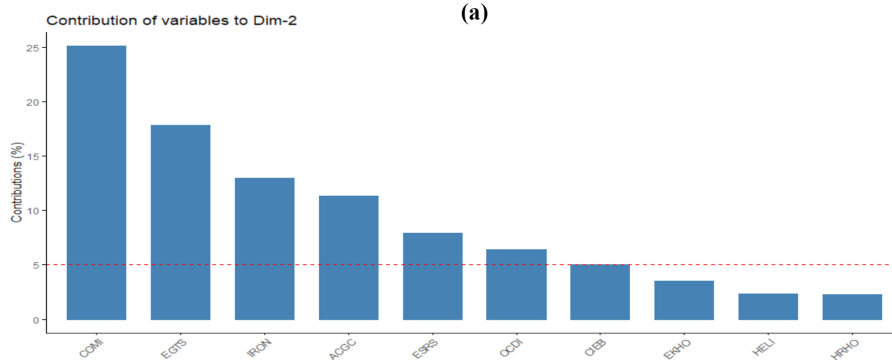


Figure 4.
The most ten
contributing stocks to
the first three PCs
during Covid-19,
respectively



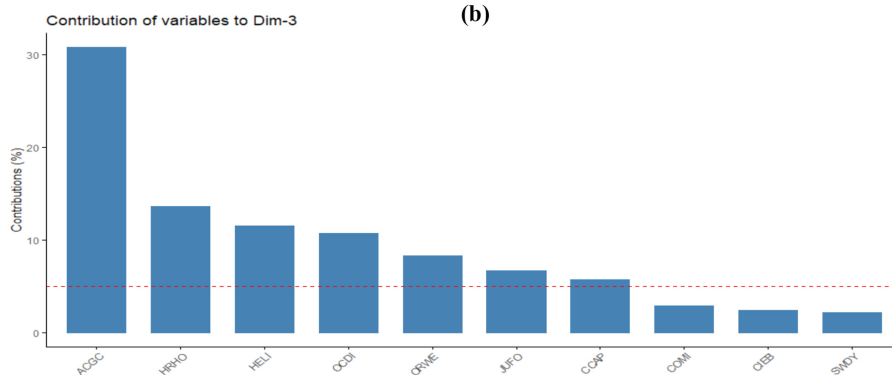
Before COVID-19

(a)



During Covid-19

(b)



Whole time-period

(c)

Figure 5.
The most ten
contributing stocks
to the first three PCs
the whole period,
respectively

the highest stocks contributing to the second PC; and ORWE, ACGC, HELI, OCDI and COMI are the highest stocks contributing to the third PC.

For the during COVID-19 time window, IRON, ESRS, EGTS, COMI, JUFO, SWDY, OCDI, EAST, MNHD and PHDC are the highest stocks contributing to the first PC; HELI, COMI, HRHO, OCDI, ORWE, EKHO and EGTS are the highest stocks contributing to the second PC; and EGCH, OCDI, ORWE, EKHO, JUFO, CCAP and ESRS are the highest stocks contributing to the third PC. For the whole-period time window, TMGH, PHDC, MNHD, SWDY, PIOH, EAST, CCAP, JUFO, CIEB and EGCH are the highest stocks contributing to the first PC; COMI, EGTS, IRON, ACGC, ESRS and OCDI are the highest stocks contributing to the second PC; and ACGC, HRHO, HELI, OCDI, ORWE, JUFO and CCAP are the highest stocks contributing to the third PC.

Linear combinations of the variables that could be conveyed by the first three PCs for the whole period could be described by the following three equations:

$$\begin{aligned} PC1 = & -0.08ACGC + 0.92CCAP + 0.88CIEB + 0.43COMI + 0.93EAST + 0.87EGCH \\ & + 0.60EGTS + 0.86EKHO + 0.82ESRS + 0.78HELI + 0.80HRHO + 0.76IRON \\ & + 0.89JUFO + 0.95MNHD + 0.71OCDI + 0.67ORWE + 0.96PHDC + 0.94PIOH \\ & + 0.95SWDY + 0.97TMGH \end{aligned} \quad (3)$$

$$\begin{aligned} PC2 = & 0.55ACGC + 0.07CCAP - 0.37CIEB - 0.81COMI + 0.07EAST - 0.13EGCH \\ & - 0.69EGTS - 0.31EKHO + 0.46ESRS - 0.25HELI - 0.25HRHO + 0.59IRON \\ & + 0.02JUFO + 0.15MNHD + 0.41OCDI + 0.11ORWE + 0.18PHDC + 0.12PIOH \\ & + 0.17SWDY - 0.03TMGH \end{aligned} \quad (4)$$

$$\begin{aligned} PC3 = & 0.70ACGC - 0.30CCAP - 0.20CIEB + 0.21COMI - 0.08EAST - 0.16EGCH \\ & + 0.14EGTS + 0.06EKHO - 0.13ESRS + 0.43HELI + 0.46HRHO - 0.06IRON \\ & - 0.32JUFO - 0.04MNHD + 0.41OCDI + 0.36ORWE - 0.02PHDC - 0.05PIOH \\ & - 0.18SWDY - 0.02TMGH \end{aligned} \quad (5)$$

By considering (3)–(5), 88% of data variability could be explained.

Table 5 reports correlation coefficients between individual stocks and the first three PCs for the before COVID-19, during Covid-19 and the whole period, respectively. Studying correlation between variables and the respective PCs is important for results interpretation (Jolliffe, 2002). Let us consider results reported in Table 5. As PC1 in all time windows explains 50% or more of the data variability, we are going to focus on analyzing PC1 in the three-time windows.

The results show that ESRS, IRON, JUFO, OCDI, PHDC and TMGH are positively related to PC1 in all the time windows. This indicates that these individual stocks had not been affected by the COVID-19 and they are moving together. This suggests that sectors of food and beverage, basic resources and real estate have not been affected by the COVID-19. Accordingly, for long-run investment decisions, investing in these individual stocks may

	Before COVID-19			During COVID-19			The whole period		
	Dim.1	Dim.2	Dim.3	Dim.1	Dim.2	Dim.3	Dim.1	Dim.2	Dim.3
ACGC	<i>0.78</i>	0.41	0.06	<i>0.91</i>	-0.20	0.18	0.30	<i>0.86</i>	-0.27
CCAP	0.25	-0.69	0.50	<i>0.85</i>	0.11	-0.32	0.49	0.25	<i>0.76</i>
CIEB	0.22	0.43	0.46	0.53	0.04	-0.09	<i>0.67</i>	-0.69	-0.16
COMI	-0.43	<i>0.76</i>	0.36	0.41	<i>0.53</i>	-0.40	<i>0.62</i>	-0.71	-0.13
EAST	<i>0.89</i>	0.24	-0.04	0.35	<i>0.78</i>	<i>0.50</i>	-0.20	<i>0.84</i>	0.35
EGCH	0.18	-0.25	<i>0.69</i>	<i>0.82</i>	-0.33	0.41	<i>0.86</i>	0.31	0.09
EGTS	-0.54	<i>0.67</i>	0.04	<i>0.96</i>	0.00	-0.14	<i>0.76</i>	-0.27	0.46
EKHO	-0.42	0.40	0.67	<i>0.96</i>	-0.06	-0.09	<i>0.74</i>	-0.28	<i>0.57</i>
ESRS	<i>0.96</i>	-0.10	-0.19	<i>0.84</i>	0.22	-0.38	<i>0.87</i>	0.04	-0.18
HELI	0.38	<i>0.83</i>	0.10	<i>0.81</i>	-0.07	0.05	<i>0.71</i>	0.40	-0.10
HRHO	0.10	<i>0.87</i>	0.24	0.17	<i>0.50</i>	-0.35	<i>0.86</i>	-0.19	-0.31
IRON	<i>0.94</i>	-0.24	-0.01	<i>0.80</i>	-0.39	0.40	<i>0.65</i>	<i>0.61</i>	-0.11
JUFO	<i>0.61</i>	-0.27	<i>0.68</i>	<i>0.94</i>	-0.08	-0.05	<i>0.76</i>	0.01	0.47
MNHD	<i>0.95</i>	0.03	-0.10	0.11	<i>0.57</i>	-0.18	0.16	<i>0.85</i>	-0.07
OCDI	<i>0.85</i>	0.40	-0.23	<i>0.98</i>	0.02	-0.12	<i>0.86</i>	-0.08	-0.08
ORWE	<i>0.57</i>	0.31	<i>0.51</i>	-0.77	0.08	0.10	<i>0.56</i>	-0.13	-0.66
PHDC	<i>0.88</i>	0.31	0.08	<i>0.98</i>	0.00	-0.07	<i>0.69</i>	0.46	-0.35
PIOH	<i>0.76</i>	-0.08	0.43	-0.16	0.38	-0.17	0.26	<i>0.69</i>	-0.41
SWDY	<i>0.91</i>	-0.10	0.08	<i>0.54</i>	<i>0.74</i>	0.39	-0.41	0.65	0.38
TMGH	<i>0.51</i>	0.43	<i>0.56</i>	<i>0.96</i>	0.05	-0.17	<i>0.88</i>	0.04	0.31

Note(s): Italic values represent the 0.01 significance level

Table 5.
Correlation between individual stocks and the first three PCs for the before COVID-19, during COVID-19 and the whole period, respectively

seem to be more stable and good decision. On the other hand, the PIOH (financial services excluding banks) and MNHD (real estate) did not play significant role during COVID-19. However, CCAP and EKHO (financial services excluding banks) EGCH (chemicals) and EGTS (travel and leisure) are positively correlated during COVID-19.

Finally, let us consider the results of implementing the PCR. As past effect cannot be ignored, the prediction model is run considering the whole period. Figure 6 shows the results of implementing cross-validation method to verify the best model. The figure displays that, the minimal Mean Squared Error of Prediction (MSEP) is approached at the first PC. Thereafter, PC1 is sufficient for predicting EGX 30 (98% of variability is explained with PC1).

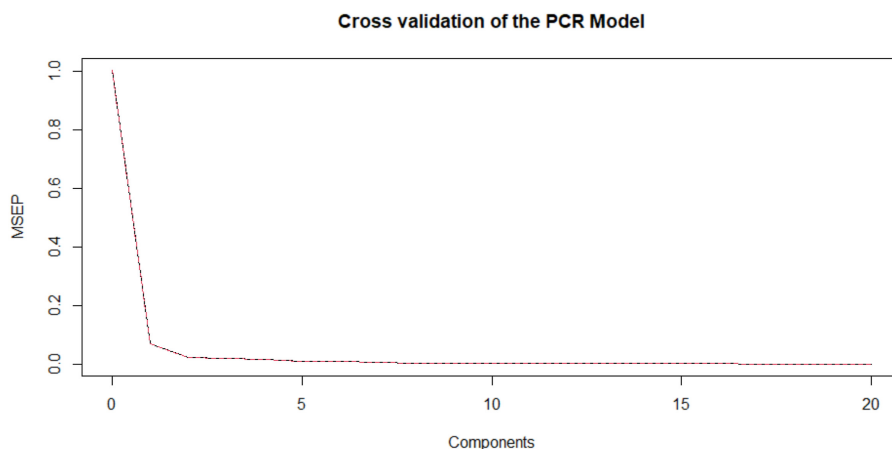
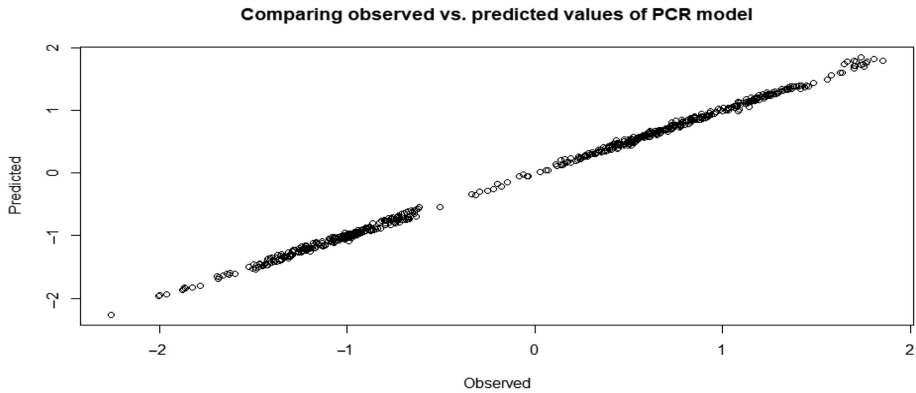


Figure 6.
Cross-validation method to identify the best model for the whole period

Figure 7.
Observed EGX
30 versus predicted
values by the PCR for
the whole period



The PCR model comprising the first PC with respect to the original predictors could be illustrated in (6) as follows:

$$\begin{aligned}
 PC1 = & -0.01ACGC + 0.07CCAP + 0.06CIEB + 0.03COMI + 0.07EAST + 0.06EGCH \\
 & + 0.04EGTS + 0.06EKHO + 0.06ESRS + 0.06HELI + 0.06HRHO + 0.05IRON \\
 & + 0.06JUFO + 0.07MNHD + 0.05OCDI + 0.05ORWE + 0.07PHDC + 0.07PIOH \\
 & + 0.07SWDY + 0.07TMGH
 \end{aligned}
 \tag{6}$$

Figure 7 depicts the observed EGX 30 values counter to the predicted values by the PCR model. The figure shows that the PCR performs very well in predicting EGX 30 values.

6. Conclusion

This paper investigates the effect of the COVID-19 on the Egyptian Stock Exchange EGX. This is achieved by applying an unsupervised machine learning method, which is the principal component analysis (PCA). PCA is a dimensionality reduction technique, which aims to find a linear combination of the original variables that explains most of the data variability. A total period of three years was studied. This period includes 16 months before the COVID-19 and 16 months during it. Three PCA models were built: (1) before the COVID-19, (2) during the COVID-19 and (3) the whole period. The study finds that for the three periods the first three Principal Components (PCs) captures about 89%, 85% and 88% of the data variability, respectively. Moreover, the results indicate that sectors of food and beverage, basic resources and real estate have not been affected by the COVID-19. Real estate stability could be explained by huge investments of the government in this sector especially the construction and development of the new capital. Policymakers could benefit from these results by issuing different regulations to support negatively affected sectors. Researchers need to study the economy features that could absorb the negative impact of the pandemic on financial markets. Moreover, a PCR model was developed to predict the future performance of the EGX 30. By comparing the observed values of EGX 30 with the predicted ones (R -squared estimated as 0.99), we can conclude that the PCR model performs fairly well.

References

- Amin, A., Arshad, M., Sultana, N. and Raoof, R. (2021), "Examination of impact of COVID-19 on stock market: evidence from American peninsula", *Journal of Economic and Administrative Sciences*, Vol. ahead-of-print No. ahead-of-print, pp. 1026-4116.
- Awad, M.H. (2020), "The impact of Covid-19 pandemic outbreak on the Egyptian firms' stock returns", *Scientific Journal for Accounting Studies*, Vol. 4, pp. 691-718.
- Baker, S.R., Bloom, N., Davis, S.J., Kost, K., Sammon, M. and Viratyosin, T. (2020), "The unprecedented stock market reaction to COVID-19", *The Review of Asset Pricing Studies*, Vol. 10, pp. 742-758.
- Burgette, L.F. and Reiter, J.P. (2010), "Multiple imputation for missing data via sequential regression trees", *American Journal of Epidemiology*, Vol. 172 No. 9, pp. 1070-1076.
- Cao, J. and Wang, J. (2020), "Exploration of stock index change prediction model based on the combination of principal component analysis and artificial neural network", *Soft Computing*, Vol. 24, pp. 7851-7860.
- Cavalcante, R.C., Brasileiro, R.C., Souza, V.L.F., Nobrega, J.P. and Oliveira, A.L.I. (2016), "Computational intelligence and financial markets: a survey and future directions", *Expert Systems with Applications*, Vol. 55, pp. 194-211.
- Cont, R. (2001), "Empirical properties of asset returns: stylized facts and statistical issues", *Quantitative Finance*, Vol. 1 No. 2, pp. 223-236.
- Contessi, S. and De Pace, P. (2021), "The international spread of COVID-19 stock market collapses", *Finance Research Letters*, Vol. 9, pp. 1-11.
- Elsayed, A. and Abdelrhim, M. (2020), *The Effect of Covid-19 Spread on Egyptian Stock Market Sectors*, SSRN, available at: <https://ssrn.com/abstract=3608734>.
- Ghorbani, M. and Chong, E. (2020), "Stock price prediction using principal component", *PLOS ONE*, Vol. 15 No. 3, pp. 1-20.
- Hargreaves, C.A. (2019), "An automated stock investment system using machine learning techniques: an application in Australia", *World Academy of Science, Engineering and Technology International Journal of Mathematical and Computational Sciences*, Vol. 13 No. 10, pp. 189-192.
- He, Q., Liu, J., Wang, S. and Yu, J. (2020), "The impact of COVID-19 on stock markets", *Economic and Political Studies*, Vol. 8 No. 3, pp. 275-288.
- Hong, H., Bian, Z. and Lee, C. (2021), "COVID-19 and instability of stock market performance: evidence from the US", *Financial Innovation*, Vol. 7 No. 12, pp. 1-18.
- Izzeldin, M., Muradoğlu, Y.G., Pappas, V. and Sivaprasad, S. (2021), "The impact of Covid-19 on G7 stock markets volatility: evidence from a ST-HAR model", *International Review of Financial Analysis*, Vol. 74 No. 101671, pp. 1-12.
- Jolliffe, I.T. (2002), *Principal Component Analysis*, 2nd ed., Springer, New York.
- Kaiser, H.F. (1960), "The application of electronic computers to factor analysis", *Educational and Psychological Measurement*, Vol. 20 No. 1, pp. 141-151.
- Liu, Z., Huynh, T.L.D. and Dai, P.-F. (2021), "The impact of COVID-19 on the stock market crash risk in China", *Research in International Business and Finance*, Vol. 57 No. 101419, pp. 1-10.
- Mazur, M., Dang, M. and Vega, M. (2021), "COVID-19 and the march 2020 stock market crash. Evidence from S&P1500", *Finance Research Letters*, Vol. 38 No. 101690, pp. 1-8.
- Mishkin, F.S. (2010), *The Economics of Money, Banking and Financial Markets*, Pearson Education, Boston.
- Sharif, A., Aloui, C. and Yarovaya, L. (2020), "COVID-19 pandemic, oil prices, stock market, geopolitical risk and policy uncertainty nexus in the US economy: fresh evidence from the wavelet-based approach", *International Review of Financial Analysis*, Vol. 70 No. 101496, pp. 1-9.

- Sharma, C. and Banerjee, K. (2015), "A study of correlations in the stock market", *Physica A: Statistical Mechanics and Its Applications*, Vol. 432, pp. 321-330.
- Solnik, B., Boucrelle, C. and Le Fur, Y. (1996), "International market correlation and volatility", *Financial Analysts Journal*, Vol. 52 No. 5, pp. 17-34.
- Uddin, M., Chowdhury, A., Anderson, K. and Chaudhuri, K. (2021), "The effect of COVID – 19 pandemic on global stock market volatility: can economic strength help to manage the uncertainty?", *Journal of Business Research*, Vol. 128, pp. 31-44.
- Waqar, M., Dawood, H., Guo, P., Shahnawaz, M.B. and Ghazanfar, M.A. (2017), "Prediction of stock market by principal component analysis", *2017 13th International Conference on Computational Intelligence and Security (CIS)*, Hong Kong, China, pp. 599-602.
- Yousfi, M., Ben Zaied, Y., Ben Cheikh, N., Lahouel, B. and Bouzgarrou, H. (2021), "Effects of the COVID-19 pandemic on the US stock market and uncertainty: a comparative assessment between the first and second waves", *Technological Forecasting and Social Change*, Vol. 167, p. 120710.
- Zhang, H.L. (2018), "The forecasting model of stock price based on PCA and BP neural network", *Journal of Financial Risk Management*, Vol. 7, pp. 369-385.
- Zhong, X. and Enke, D. (2019), "Predicting the daily return direction of the stock market using hybrid machine learning algorithms", *Financial Innovation*, Vol. 24 No. 5, pp. 1-20.

Further reading

- Baba, Y. and Nakamura, T. (2002), "Time dependent principal component analysis", in Nishisato, S., Baba, Y., Bozdogan, H. and Kanefuji, K. (Eds), *Nishisato Measurement and Multivariate Analysis*, Springer, Tokyo.
- Chen, S. and Yeh, C. (2002), "On the emergent properties of artificial stock markets: the efficient market hypothesis and the rational expectations hypothesis", *Journal of Economic Behavior and Organization*, Vol. 49 No. 2, pp. 217-239.
- Fama, E.F. (1970), "Efficient capital markets: a review of theory and empirical work", *The Journal of Finance*, Vol. 25 No. 2, pp. 383-417.
- Hanspal, T., Weber, A. and Wohlfart, J. (2020), "Exposure to the COVID-19 stock market crash and its effect on household expectations", *The Review of Economics and Statistics*, Vol. 103, pp. 1-45.

Corresponding author

Heba M. Ezzat can be contacted at: hebaezzat@cu.edu.eg