RESEARCH ARTICLE
https://doi.org/10.17059/ekon.reg.2022-4-24
UDC 336.7

Dana Al-Najjar [a], Hazem Al-Najjar [b], Nadia Al-Rousan [c]
[a] Applied Science Private University, Amman, Jordan
[b] Istanbul Gelisim University, Istanbul, Turkey
[c] Sohar University, Sohar, Oman

LONG-TERM GENERAL INDEX PREDICTION BASED ON FEATURE SELECTION AND SEARCH METHODS: AMMAN STOCK EXCHANGE MARKET[1]

Abstract. Stock markets are an essential backbone for the economy worldwide; their indices provide all interested parties with indicators regarding the performance of firms listed in the financial market due to tracking the daily transactions. This study aims to investigate factors that affect the stock exchange directly so that it simplifies building a prediction model for the exchange index in Jordan's financial market. The study hypothesis assumes that some sub-sectors are most influential in creating the stock market prediction model. Therefore, this study applies four feature selection methods on 23 sub-sectors and Amman Stock Exchange Index (ASEI100) for the period 2008–2018. The top 10 attributes from each selection method are combined, and the frequency table is used to find the highly trusted attributes. Moreover, linear regression with ordinary least square regression is used to test the validity of the top factors that frequently occurred in the four methods and their effect on ASEI. The results found that there are six main sub-sectors directly affecting the general index in Jordan: Health Care Services, Mining and Extraction Industries, Textiles, Leather and Clothing, Real Estate, Financial Services and Transportation. These sectors can be utilised to predict the movements of the Amman Stock Exchange Index in Jordan. Also, the linear regression model output showed a statistically significant relationship between the six sub-sectors (independent variables) and ASEI (dependent variable). Investors can use this paper's findings to signal the most important sectors in Jordan. Thus, it helps in taking investment decisions.

Keywords: Amman stock index, feature selection and search methods, linear regression, economic sectors, prediction models, financial services, long-term general index, Syrian refugees, Syrian War, correlation analysis

For citation: Al-Najjar, D., Al-Najjar, H. & Al-Rousan, N. (2022). Long-Term General Index Prediction Based on Feature Selection and Search Methods: Amman Stock Exchange Market. *Ekonomika regiona / Economy of regions, 18(4)*, 1301-1316, https://doi.org/10.17059/ekon.reg.2022-4-24.

[1] © Al-Najjar D., Al-Najjar H., Al-Rousan N. Text. 2022.

ИССЛЕДОВАТЕЛЬСКАЯ СТАТЬЯ

Д. Аль-Наджар [а], *Н. Аль-Наджар* [б] ✉, *Н. Аль-Роусан* [в]

[а] Частный университет прикладных наук, г. Амман, Иордания
[б] Стамбульский университет Гелишим, г. Стамбул, Турция
[в] Сохарский университет, г. Сохар, Оман

Применение метода отбора признаков для долгосрочного прогноза индекса Амманской фондовой биржи

Аннотация. Фондовые биржи — неотъемлемая часть мировой экономики; благодаря отслеживанию ежедневных операций, фондовые индексы отражают изменения показателей деятельности представленных на финансовом рынке фирм. Для построения модели прогнозирования фондового индекса Иордании в данной статье исследованы факторы, напрямую влияющие на индекс фондовой биржи. Чтобы выявить, какие секторы экономики оказывают наибольшее влияние на модель прогнозирования, авторы применили четыре метода отбора признаков для изучения связи между 23 секторами и индексом Амманской фондовой биржи (ASEI100) за период 2008–2018 гг. В каждой модели были выделены 10 наиболее значимых факторов, которые затем они были объединены и внесены в таблицу частот. Для проверки достоверности основных факторов, которые наиболее часто встречались в четырех моделях, а также для оценки их влияния на ASEI использовались методы линейной регрессии и обычных наименьших квадратов. Результаты исследования показали, что существует шесть основных секторов, непосредственно влияющих на общий фондовый индекс в Иордании: здравоохранение, горнодобывающая промышленность, производство одежды, текстиля и изделий из кожи, недвижимость, финансовые услуги, транспорт. Показатели этих секторов можно использовать для прогнозирования изменений индекса Амманской фондовой биржи в Иордании. Кроме того, линейная регрессия выявила статистически значимую взаимосвязь между шестью секторами (независимые переменные) и ASEI (зависимая переменная). Полученные результаты, описывающие наиболее важные секторы экономики Иордании, могут быть использованы инвесторами для принятия инвестиционных решений.

Ключевые слова: фондовый индекс Аммана, методы отбора признаков, линейная регрессия, секторы экономики, модели прогнозирования, финансовые услуги, долгосрочный индекс, сирийские беженцы, война в Сирии, корреляционный анализ

Для цитирования: Аль-Наджар Д., Аль-Наджар Н., Аль-Роусан Н. (2022). Применение метода отбора признаков для определения влияния секторов экономики на долгосрочный прогноз индекса Амманской фондовой биржи. *Экономика региона, 18(4)*, 1301-1316. https://doi.org/10.17059/ekon.reg.2022-4-24.

1. Introduction

Stock market analysis has always been an interesting topic for researchers and investors because of the highly interrelated macro and micro factors, which include economics, politics, psychology and company-specific aspects. Therefore, over the years, interested parties have come up with various theoretical foundations in mathematics that facilitate the development of various methods and models for stock market analysis; besides, it enables the prediction of stock prices using various mathematical methods.

Multiple theories have contradictory outcomes regarding the ability to predict stock prices by applying historical data. One of the most influential theories was the Efficient Market Hypothesis (EMH) proposed by Fama in 1970, which states that stock prediction cannot be achieved using historical data because all information is already reflected in stock prices. On the other hand, although the efficient market hypothesis (EMH) opposes predicting stock prices, especially with samples from emerging markets, for many reasons, such markets are inefficient at all levels (Lo, Mackinaly, 1988; Cao, Parry, Leggio, 2011).

Moreover, stock prediction is a vital and challenging issue in finance because of the extensive database of prices that managers, analysts, and other parties need to work through to classify and predict them. That all sheds light on moving to the next challenging stage: depending on sophisticated computer programmes to manage the available data through data mining mechanisms to facilitate prediction. One method that proves to be efficient in this matter is called the feature selection method. It is used to filter redundant and/or irrelevant features. In addition, it evaluates many significant factors in the stock market to depict the optimal feature subset that provides the appropriate predictive power in modelling the available data.

This study aims to answer many questions. Which are the most critical sub-sectors in the Amman stock exchange that can affect a gen-

eral index? Can a feature selection be used to select the most connected sub-sectors to the general index? Can a linear regression model be used to validate the capability of stock index prediction? This study used a long-term dataset from Amman Stock exchange indices for ten years (2008–2018). After selecting the optimal attributes, a linear regression model was applied to verify the sectors' capability to estimate a general index and accurately predict stock market prices and indices.

The remainder of this paper is organised as follows. Section 2 contains previous studies and a theoretical framework, and Section 3 explains the case study of the Amman stock exchange market and its methodology. Section 4 demonstrates the method adopted in this paper. Section 5 summarises the main descriptive statistical tests. Section 6 illustrates the feature selection methods applied in this study and their outputs, also the outcome of the Ordinary Least Square methods. Section 7 summarises the conclusions of the tests used in this paper.

2. Previous Studies and Theoretical Framework

The excellent performance of stock markets is essential for the healthiness of the country's economy, acting as a good signal for economic growth. The general index is one of the most important indicators of the stock market's performance. The general index of the Stock Exchange tracks the movements of the buying and selling prices of companies' shares in the market. Also, it can refer to the market's purchasing power in general. The general index is defined as the weighted index of market capitalisation, and it is calculated by considering the most important and prominent companies in the market.

On the other hand, sub-sectors are measured according to the firms included in each sector; for example, the banks' index is calculated by utilising the prices of bank shares listed at the Amman Stock exchange containing representative banks in the financial market. Many studies have focused on explaining the trend and behaviour of the market using only the sub-sectors without considering all variables that were found to be related to previous studies. That is due to restrictions attached to higher cost, time-consuming, and complicated process with various probabilities distributions for the selected variables, which affects the conclusions negatively. Besides that, various scholars and researchers investigated many statistical tests to choose the most influential variables that directly affect the financial markets by adopting different parameters (such as the general stock exchange index, stock prices, oil prices, firms' bankruptcy and so forth). Thus, it facilitates decision-making by investment managers and interested parties regarding both finances and investment. Across the years, numerous parameters were applied to specify the significant variables of the financial market. Chundakkadan and Sasisdharan (2019) studied the relationship between the Central Bank's stock returns and daily open market operations (OMO). To make the study unique, the authors used two new monetary policy tools, Repo Spread and Reverse Repo Spread. The results showed that the Central Bank's money market operations significantly affect daily stock returns; also, the prediction of stock returns can be improved by using monetary policy variables. Also, Choudhry and Osoble (2015) investigated long-run and short-run relationships between the United States' stock industrial sectors and three emerging markets, including Brazil, Malaysia, and South Africa. After analysing data using different nonlinear econometrics time series techniques, the results of the time series analysis revealed that global financial markets tend to be volatile and turbulent. The authors explained that the conclusion of their study could motivate investors to focus on investing in specific industrial sectors. Syed Nor, Ismail and Yap (2019) proposed a personal bankruptcy prediction model build an alert system for the Malaysian economy to minimise personal bankruptcy cases since the total personal bankruptcy cases in 2014 reached 131,282 in Malaysia. The prediction system is based on evaluating 12 predictors and choosing the most efficient model among overall models. After analysing different prediction models, the results showed that using all the variables may not improve the prediction model and selecting the most related will enhance the prediction.

Furthermore, Gupta et al. (2019) investigated the ability to predict West Texas Intermediate oil returns and volatility by using news-based measure financial stress index (FSI). Dynamic conditional correlation multivariate generalised autoregressive conditional heteroscedasticity (DCC-MGARCH) was used to validate the relation between the two variables. The results showed that the linear Granger test failed to detect any evidence of predictability. The results revealed that using FSI as an independent variable to predict oil returns is suitable for using a nonlinear relationship.

Bhuiyan et al. (2019) investigated the relationship between Sukuk (Islamic bond) and global diversification and analysed the advantages that are expected from selling Sukuk by using wave-

let coherence and multivariate-GARCH analyses. The study considered the volatilities and correlations of sovereign bond indexes in different countries, including the USA, Canada, Germany, the UK, Australia, and Japan, and the Thomson Reuters BPA Malaysia Sukuk Index. The results showed that the Malaysian Sukuk index has a lower co-movement than the developed market bond index returns. Besides, the US and Canadian bond markets have a negative uncorrelated relationship with the Malaysian Sukuk market.

However, Kirikkaleli (2018) investigated the effect of Taiwan's domestic and foreign factors on the stock market index by considering different tests, including autoregressive distributed lag (ARDL), Dynamic Ordinary Least Squares Estimator (DOLS) and Markov Switching. The long-term test results found that combining two factors, mainly domestic and foreign, strongly affects the stock market index. Also, Moghaddam, Moghaddam and Esfandyari (2016) investigated the ability to use different feed forward back propagation artificial neural networks in forecasting the daily stock exchange index of NASDAQ. Short-term historical stock prices are considered for six months to build a prediction model. The results found that neural network can remarkably forecast NASDAQ index.

On the other hand, more studies were interested in significant financial market indices, such as Qiu and Song, (2016), which investigated the ability to predict the next day's price of the Japanese stock market index using two input variables. An optimised neural network using a genetic algorithm is proposed to predict the market indices. Compared with prior studies, the results revealed that choosing the best variables to build a prediction model can improve the performance of predicting the stock index market in Japan. Also, Hu et al. (2018) proposed an improved neural network to predict the opening of S&P 500 and Dow Jones Industrial stock indices. The enhanced neural network is based on the sine and cosine algorithm, built on adjusting the weights and basis of back propagation neural networks. Two scenarios are considered to validate the prediction system, including prediction without Google Trends and with Google Trends. The results showed that the performance of the proposed neural network is higher than prior studies that used other optimisations like Particle swarm optimisation. Finally, after comparing all the models and analysing their performance, the results revealed that using Google trends can improve the performance of the prediction system. Chen and Hao (2018) investigated Shanghai and Shenzhen stock exchange markets prediction by integrating Principal Component Analysis (PCA) and Weighted Support Vector Machine (WSVM), besides combining different integrations between PCA and other prediction models. The results showed that the proposed model is better than other models in predicting the stock market movement.

Moreover, many researchers have been working on understanding and investigating the stock market's performance by predicting stock prices and general index movements. Various machine learning models were used in the methodologies, including multiple neural networks (Chauhan, Ravi, Chandra, 2009; De Faria et al., 2009; Adebiyi, Adewumi, Ayo, 2014; Zahedi, Rounaghi, 2015; Rather, Agarwal, Sastry, 2015; Inthachot, Boonjing, Intakosum, 2016; Khuat et al., 2016; Chandar, Sumathi, Sivanandam, 2016; Wang et al., 2016; Hiransha et al., 2018; Chen, Hao 2018; Zhou et al., 2019; Orimoloye et al., 2020; Al-Najjar D. et al., 2022; Al-Najjar H. et al., 2021), Support Vector Machines (Nahil, Lyhyaoui, 2018; Tang, Dong, Shi, 2019; Umer, Sevil, Sevil, 2019), Autoregressive Integrated Moving Average (ARIMA) (Umer, Sevil, Sevil, 2019; Assous et al., 2020). Moreover, some researchers proved that linear regression models could outperform the performance of machine learning models and improve the linear estimation for independent variables, as discussed by Al-Najjar D. et al. (2022).

According to the previously mentioned studies, we can conclude that there is no optimal model that can be fitted for all stock market indices worldwide. Hence, before building a prediction model for a stock market, it is essential to analyse data to check and define the important variables that need to be applied to improve the prediction model. Thus, determining the significant variables directly related to predicting stock market prices and indices can help overcome biased or incorrect variables. This study is interested in applying the Feature Selection Method on Amman Stock Exchange Index (ASEI100) to select the sub-indices that are directly and significantly related to the index during the period of (2008–2018) to facilitate building the prediction model for future studies.

3. Case Study: Amman Stock Exchange

Jordan is one of the Middle Eastern countries that is located in Western Asia on the East Bank of the Jordan River. Jordan is bordered by four countries, including Saudi Arabia (south and east), Iraq (northeast), Syria (north) and Palestine (west). Its economy can be described as an emerging market upper-middle-income country, as mentioned

by the World Bank. In the year 2000, Jordan became the first Arab country to join the World Trade Organisation (WTO), which allowed Jordan to establish free trade with the United States; besides that, Jordan got a high capability to export its products to European markets too. In addition, GDP between 2004 and 2008 grew at an average rate of 8 % annually and then dropped to 2.6 % from 2010 onwards until now. In 2011, the Arab Spring caused recession and turmoil in many Arab countries besides neighbour countries. This decreased Jordan's GDP growth and negatively affected many sectors of the Jordanian economy, including trade, industry, construction and tourism. Due to the region's unstable political conditions, many obstacles occurred, such as the decrease in both tourist levels and foreign investments that in return led to an increase in both military expenditures, and hosting refugees' cost.

Jordan's economy depends on various sectors, including transportation and communication, public utilities, construction accounts, mining and manufacturing constitute; thus, it is a well-diversified economy. Furthermore, the stock markets were established to provide capital to firms and to allow sharing the firm's profits among many investors. Amman Stock Exchange (ASE) was established in 1999 as a regulated market for trading securities in Jordan. In 2017, ASE was registered as a public shareholding company. ASE's main duties include operating, managing and developing the operation markets inside and outside Jordan.

4. General Index Calculation

This study applies the feature analysis on the index of Jordan's stock exchange market, the Amman Stock Exchange Index ($ASEI_{100}$). The calculation of the index depends on the most liquid and largest firms in the stock market, and its main aim is to show how the stock market prices fluctuate up and down during the trading sessions. Interested parties such as investors and financial managers depend on the general index to understand the market trend to predict the movements of the stock market, which will help them make accurate investment decisions, including issuing, buying, or selling.

The stock market index is a hypothetical portfolio of a firm's stocks listed on the stock exchange representing a market segment. One of the main features of ASEI is that it depends on the free float. Hence, it gives a better reflection of the changes in stock prices and is less biased toward large market capitalisation, which in return provides more support to the prediction power of the model depending on the general index.

The Amman Stock Exchange Index (ASEI) consists of 100 of the most liquid and largest firms listed at the Jordanian Stock Exchange. It is calculated using the Paasche method that is adopted by various exchanges worldwide as follows:

$$Index\ (t) = (M_t / B_t) \cdot 1000, \qquad (1)$$

where M_t is the market capitalisation of constituents at time t (the sum of the market capitalisation of all stocks included in the index), B_t is the base value of the index. The process consists of three steps which are:

1. For each selected firm, the market capitalisation will be calculated by multiplying the market price by the number of outstanding shares issued by the company.

2. The free-float market capitalisation of each firm is calculated by multiplying the market capitalisation with the free-float factor that lies between 0.05 and 1.00. The equation to calculate the free-float factor is the shares owned by the board of directors and investors who own more than 5 % of the shares that are subtracted from the total listed shares.

3. The free-float market capitalisation is divided by a number known as the Index Base. Amman stock exchange (ASE) used an index base of 100 points on December 31, 1991 and the index base value was changed to 1000 as of January 1, 2004, based on rules for the ASE weighted index.

5. Research Methodology

This section introduces the methodology applied in the study. Most vital stock market decisions are focused on investing or financing. The firm can fulfil its main objectives and continue to exist if these decisions are accurate. Therefore, it is crucial to predict stock prices to facilitate decision-making and reduce the time and cost required for efficient management. The Amman Stock Exchange Index (ASEI) represents Jordan's stock market and is our sample. In addition, the applied method used a feature selection to identify the major attributes related directly to the targeted outcome, which is ASEI.

It is well known that no optimal model can match all stock market indices worldwide. Therefore, to build a prediction model for any stock market, the data must be analysed to check the most significant variables that need to be applied to improve the prediction. Hence, this paper collected the data of stock sub-sector indices and the general index from the Jordanian financial market for the years 2008 and 2018. The first step is to analyse and clean the data to ensure no outlier data and no null values are present in

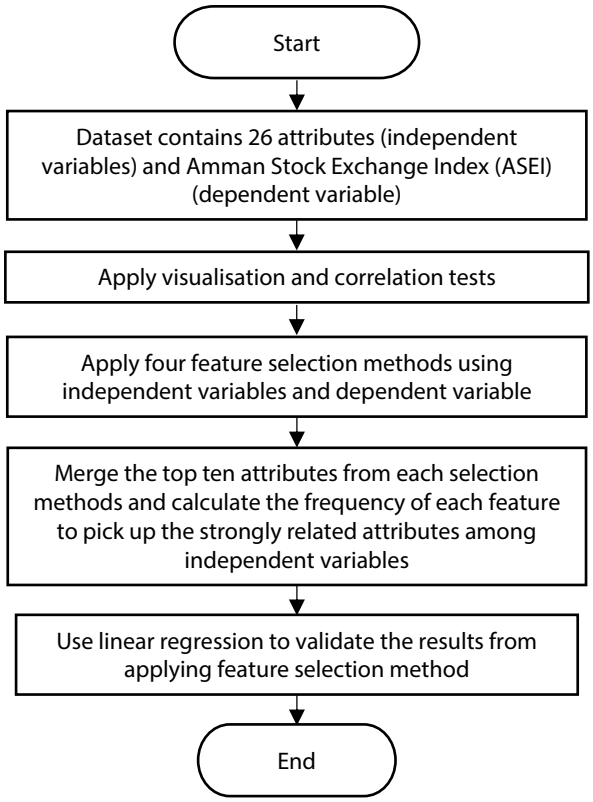


Fig. 1. *Steps of building feature selection method*

the collected data. Then it is necessary to visualise the graphs of all sub-sectors (the independent variables) against the General Stock Market Index (ASEI) (the dependent variable), as well as find their correlation between the variables. Thus, the clean data is ready to be applied to four feature selection methods to identify the most important independent variables (i.e., sector indices). Then, the top ten independent variables from each of the four attribute selection methods will be chosen to construct a model that contains the most trusted attributes, which can enhance building a model for prediction. Finally, a multiple regression model is utilised to validate the results of the feature selection methods in defining the main attributes affecting the general stock exchange index (ASEI). The dependent variable is the general index, and the independent variables are the best attributes retrieved from the proposed feature selection methods. The flowchart for the development of the feature selection method for the proposed model is shown in Figure 1.

5.1. Attribute Selection Methods

Independent and dependent variables are two major components of any dataset. Researchers have been using various methods with different levels of sophistication, such as mathematical, statistical, and neural network, to explain the behaviour of dependent variables through the movements of independent variables. Therefore, it is essential to adopt a preliminary process to choose the relevant and significant independent variables from the whole list of variables to enhance the prediction performance and identify more representative features to be utilised. Thus, the feature selection method is one of the reliable processes for picking the optimal variables that can be applied in predicting dependent variable(s). In addition, it is a preprocessing step of data mining that can filter redundant and/or irrelevant features. In addition, it is considered to have adequate predictive power in modelling because it is a dimensionality reduction technique that selects only a subset of measured features after evaluating them to build up a representative model. Hence, it is very useful in cases of high-dimensional data, and it is convenient to be applied to stock exchange indices.

Thus, this study is interested in applying feature selection methods through selecting attributes prominent to Amman stock exchange general index (ASEI). Afterwards, they will be evaluated, the most significant factors to depict the financial market trend will be determined. Feature Selection is divided into two stages: attribute evaluator and search methods. Firstly, the attribute evaluator uses each attribute (variable) in the dataset to evaluate the selected dependent variable; it contains many algorithms, including correlation-based feature (Hall, 1999), correlation attribute evaluator, relief attribute evaluator (Kononenko, 1994), and classifier attribute evaluator. So, our study is interested in applying four algorithms encompasses: subset evaluator, correlation evaluator, relief evaluator, and classifier evaluator. Besides that, search methods will be picked to navigate different combinations of attributes to reach a short list of chosen features. Therefore, two methods will be applied, including the best first that uses greedy hill climbing augmented with a backtracking facility and ranker search method that uses conjunction with attribute evaluators. Moreover, the chosen techniques in this study for both attribute evaluator and search methods are considered one of the best combinations evaluated by many researchers in the field (Ruiz, Riquelme, Aguilar-Ruiz, 2005; Chetty, Vaisla, Sudarsan, 2015; Onik et al., 2015; Al-Najjar, D. et al., 2022).

The second part covers selecting a suitable combination of both search methods and attribute selection models, relying on the findings of previous studies to construct a solid model with the best attributes. Regarding the attribute selection method, two methods are engaged in this

Table 1

Selected Feature Selection Methods

Models	Search method	Attribute Selection
Model 1	Best First	subset evaluator
Model 2	ranker	correlation attribute evaluator
Model 3	ranker	relief attribute evaluator
Model 4	ranker	classifier attribute evaluator

study, including the best first and ranker methods. Moreover, four methods for attribute selection are applied: subset evaluator, correlation attribute evaluator, relief attribute evaluator, and classifier attribute evaluator.

Furthermore, this study developed four models, as shown in Table 1; each model ranks all the attributes in a descending form from the highest to the lowest effect on Amman Stock Exchange Index (ASEI). The top ten outputs of each of the four selection models are picked up to be combined to prepare a frequency table. In the end, we will choose the most effective related variables that can be used to predict the general index.

After selecting the most effective sub-sectors that can efficiently affect a general index in Jordan, a linear regression model is used to verify the selected sub-sectors. Besides, the significance and the variable coefficients are considered to check the direction and a linear weight for each independent variable.

Table 2

Attributes statistical description

Attribute / Sectors	Min	Max	Mean	Std. Deviation
Day	1	31	16	9
Month	1	12	6	3
Year	2008	2018	2013	3
Banks	3264	5824	4060	520
Insurance	1718	4375	2309	570
Financial Services	1255	10119	2757	2051
Real Estate	1318	9031	2687	1674
Health Care Services	735	1231	889	75
Educational Services	2366	4231	3170	519
Hotels and Tourism	950	2507	1390	381
Transportation	287	1546	584	295
Technology and Communications	384	2766	1350	588
Media	52	4492	1252	1222
Utilities and Energy	2891	10786	4467	1015
Commercial Services	724	2495	1209	332
Pharmaceutical and Medical Industries	795	2587	1697	405
Chemical Industries	1153	2440	1567	230
Paper and Cartoon Industries	57	1419	320	351
Printing and Packaging	1266	6186	2772	1512
Food and Beverages	1060	2696	1521	371
Tobacco	1456	27517	8702	7146
Mining and Extraction Industries	1287	11549	3108	1786
Engineering and Construction	712	3869	1835	684
Electric Industries	944	10667	2818	2300
Textiles, Leather and Clothing	1519	2914	1929	242
Glass and Ceramic Industries	438	2102	1101	256
General Index	1801	5044	2337	582

6. Results, Analysis and Discussion

6.1. Description Analysis

Before applying the feature selection algorithms, it is essential to go through various preliminary statistical tests to study the trend and features of the selected data. These tests include statistical description, correlation, and visualisation. The study sample consists of the Amman Stock Exchange Index (ASEI100), around 23 sub-sector indices with the day, month, and year of the trading session covering the period of 2008–2018. Table 2 presents a major statistical description which includes Max, Min, Mean, and standard deviation for all sub-sectors indices and the Amman Stock Exchange General Index (ASEI) for the period of 11 years (2008–2018)

The result of the statistical description can be summarised as follows. In terms of column minimums, it was lowest in both the media and paper and board industries. Besides, the highest variables were tobacco, mining and extraction, the electric industry, utility and energy, and financial services. As for the mean, the highest is for tobacco. Moreover, the last column is for standard deviation in which tobacco, electric industries, financial services and real estate have the widest variables among its data.

Table 3 demonstrates the correlation between all variables and the Amman stock exchange index (ASEI). One variable was found to have an insignificant relationship, which is the day. The other variables can be divided into three groups; the first group includes variables with low correlation values of less than 0.40, which are month, health care services, educational services, printing and packaging, tobacco, also glass and ceramic industries formulating around 24 % of all variables. The second group of variables with moderate correlation values between 0.50 and 0.70 are year, banks, technology and communi-

Table 3

Correlation analysis between financial variables and the general index

Attribute	Correlation coefficient	Significant (2-tailed)
Day	–0.01	0.555
Month	–0.12	0
Year	–0.60	0
Banks	0.69	0
Insurance	0.92	0
Financial Services	0.97	0
Real Estate	0.98	0
Health Care Services	0.36	0
Educational Services	–0.13	0
Hotels and Tourism	0.82	0
Transportation	0.91	0
Technology and Communications	0.61	0
Media	0.78	0
Utilities and Energy	0.82	0
Commercial Services	0.83	0
Pharmaceutical and Medical Industries	0.59	0
Chemical Industries	0.55	0
Paper and Cartoon Industries	0.89	0
Printing and Packaging	–0.23	0
Food and Beverages	0.90	0
Tobacco	–0.37	0
Mining and Extraction Industries	0.84	0
Engineering and Construction	0.74	0
Electric Industries	0.939	0
Textiles, Leather and Clothing	0.877	0
Glass and Ceramic Industries	0.251	0

cations, pharmaceutical and medical industries, and chemical industries consisting of around 20 % of all features.

The third group contains the highest percentage of all variables, around 56 % of all selected attributes have a strong correlation above 0.70, indicating that most variables are directly related. These sectors are insurance, financial services, real estate, hotels and tourism, transportation, media, utilities and energy, commercial services, paper and cartoon industries, food and beverages, mining and extraction, engineering and construction, electric industries, and textiles leather and clothing.

Before running the feature selection method, it is also beneficial to depict the relationship between each attribute selected in this study with the Amman Stock Exchange General Index (ASEI100) to visualise the direction and shape of the relationship. The relationship is visualised in Figures 2 to 4.

The visualisation results showed that eleven attributes (mainly banks, insurance, financial services, real estate, hotels and tourism, transportation, technology and communications, media, food and beverages, mining and extraction industries, engineering and construction, and electric industries) have a linear relationship with the Amman Stock Exchange Index (ASEI), where the remaining fourteen attributes have a nonlinear relationship with ASEI.

All the collected results confirmed that the selected variables have a direct relationship and influence on the ASEI index. Until this step, the study found all the attributes associated with ASEI; as a second step to validate the most important attributes of ASEI and save time and cost of ASEI consumption, four feature selection methods are applied and validated in the next section.

6.2. Applying Feature Selection Methods

This section concentrates on the outcomes of running each of the four models in Table 3. Each of the four models will rank all the selected attributes according to their direct effect on the Amman Stock Exchange Index (ASEI), specifying the essential attributes that will facilitate building a prediction model for ASEI.

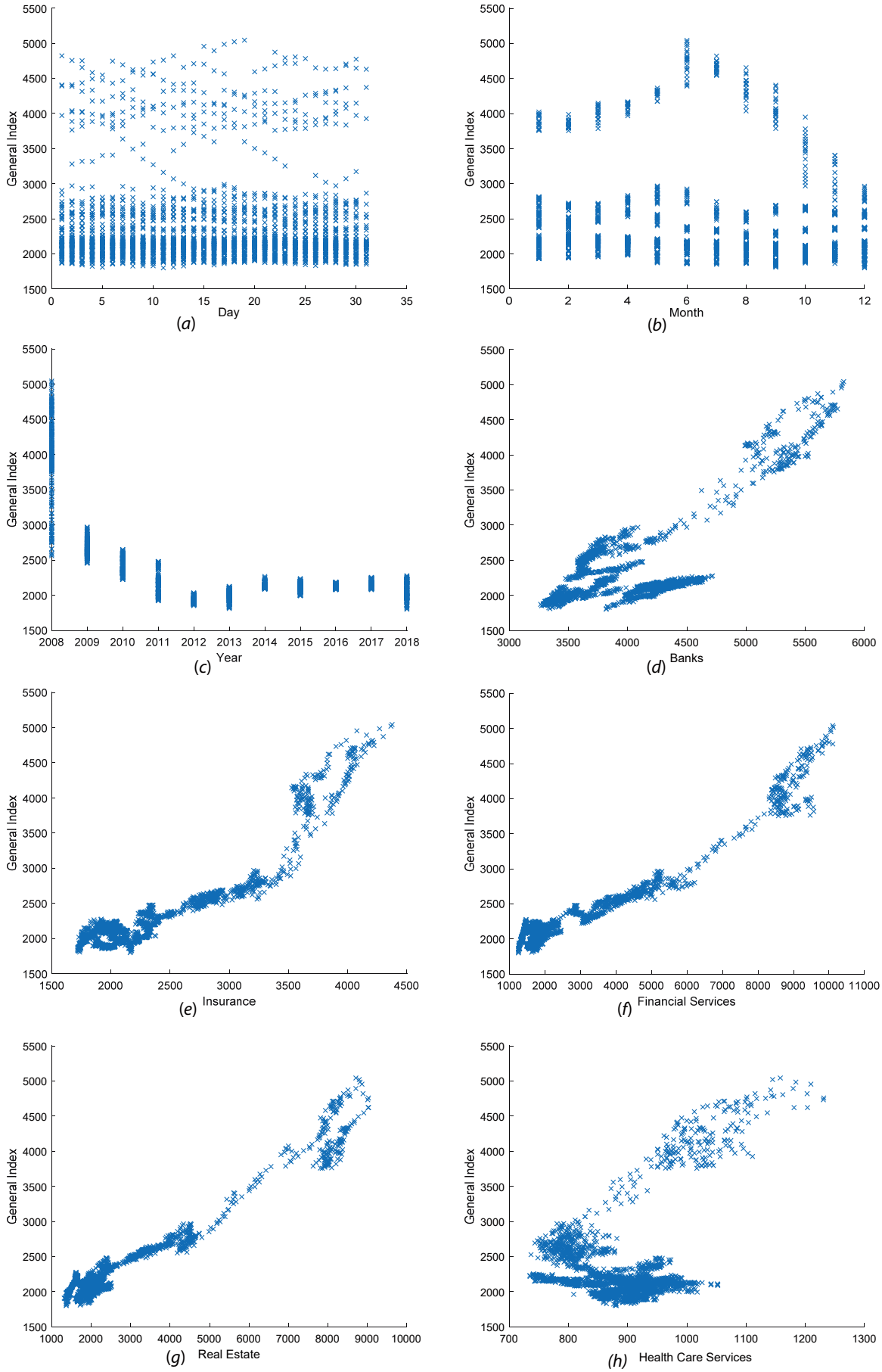
Table 4 represents the first model that combines the best first and subset evaluator. The main findings show that six main attributes are directly related to the ASEI: Real Estate, Health Care Industries, Commercial Services, Food and Beverages, Mining and Extraction Industries, and Textile, Leather and Clothing.

On the other hand, Table 5 provides the output of the second model that merges the ranker and correlation attribute evaluator. The attributes are divided into groups according to their rank and are arranged in a descending order from the highest to the lowest effect; the sectors, including Real Estate, Financial Services, Electric Industries, Insurance and Transportation, are ranked more than 0.9. Seven sectors: Food and Beverages, Paper and Cartoon Industries, Textiles, Leather and Clothing, Mining and Extraction Industries, Commercial Services,

Table 4

Results of proposed Model 1

Attribute Selected
Real Estate
Health Care Services
Commercial Services
Food and Beverages
Mining and Extraction Industries
Textiles, Leather and Clothing



The end of Figure 2 on next page

The end of Figure 2

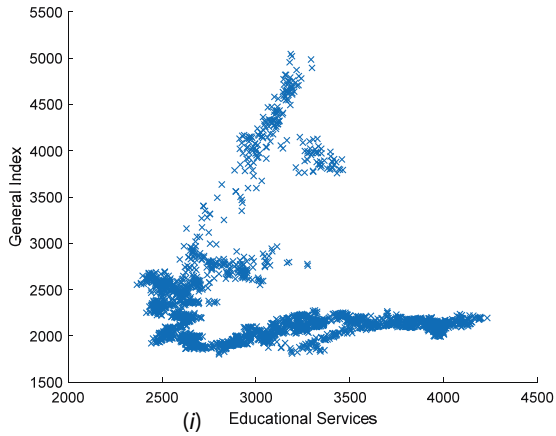
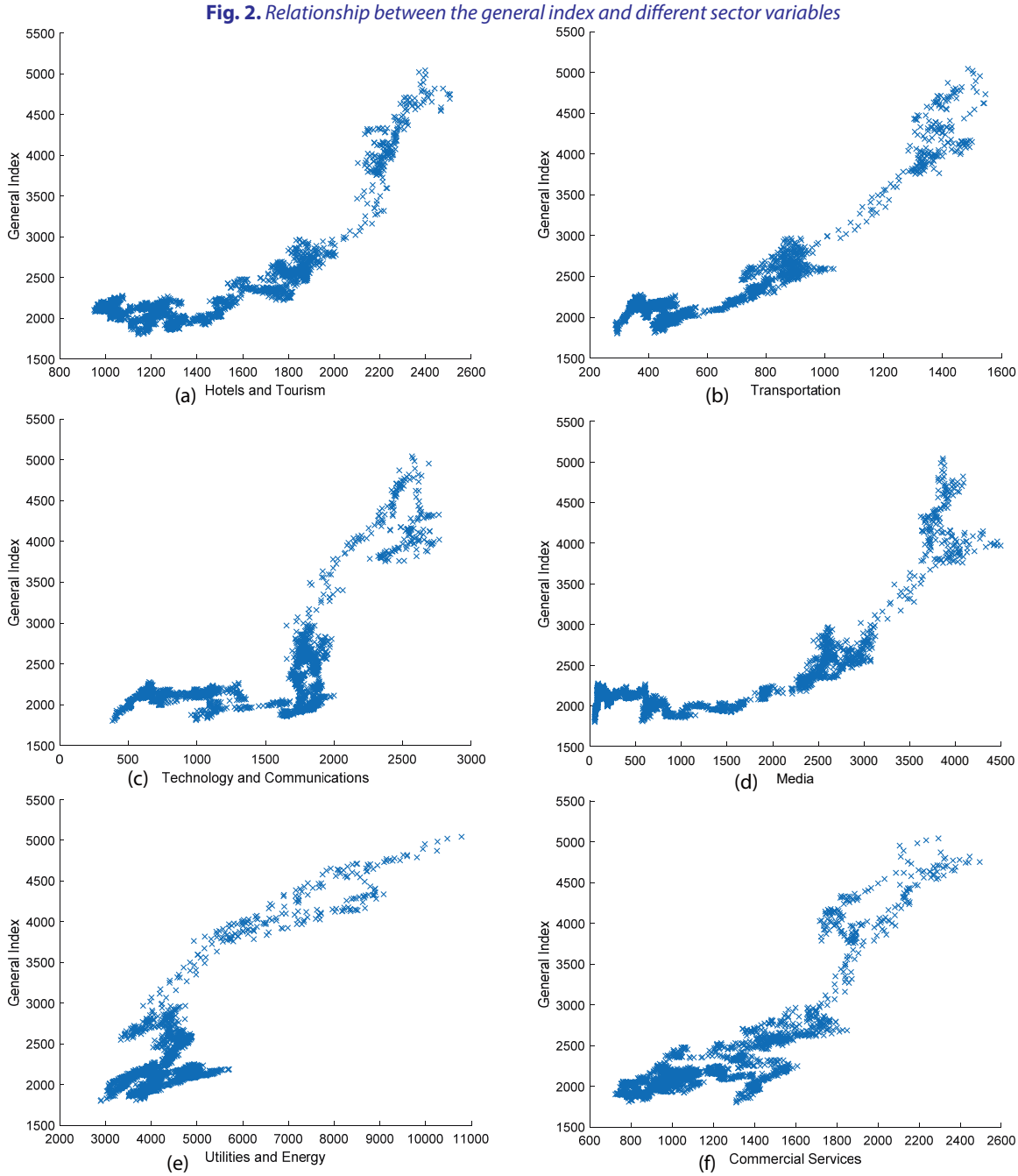


Fig. 2. *Relationship between the general index and different sector variables*



The end of Figure 3 on next page

The end of Figure 3

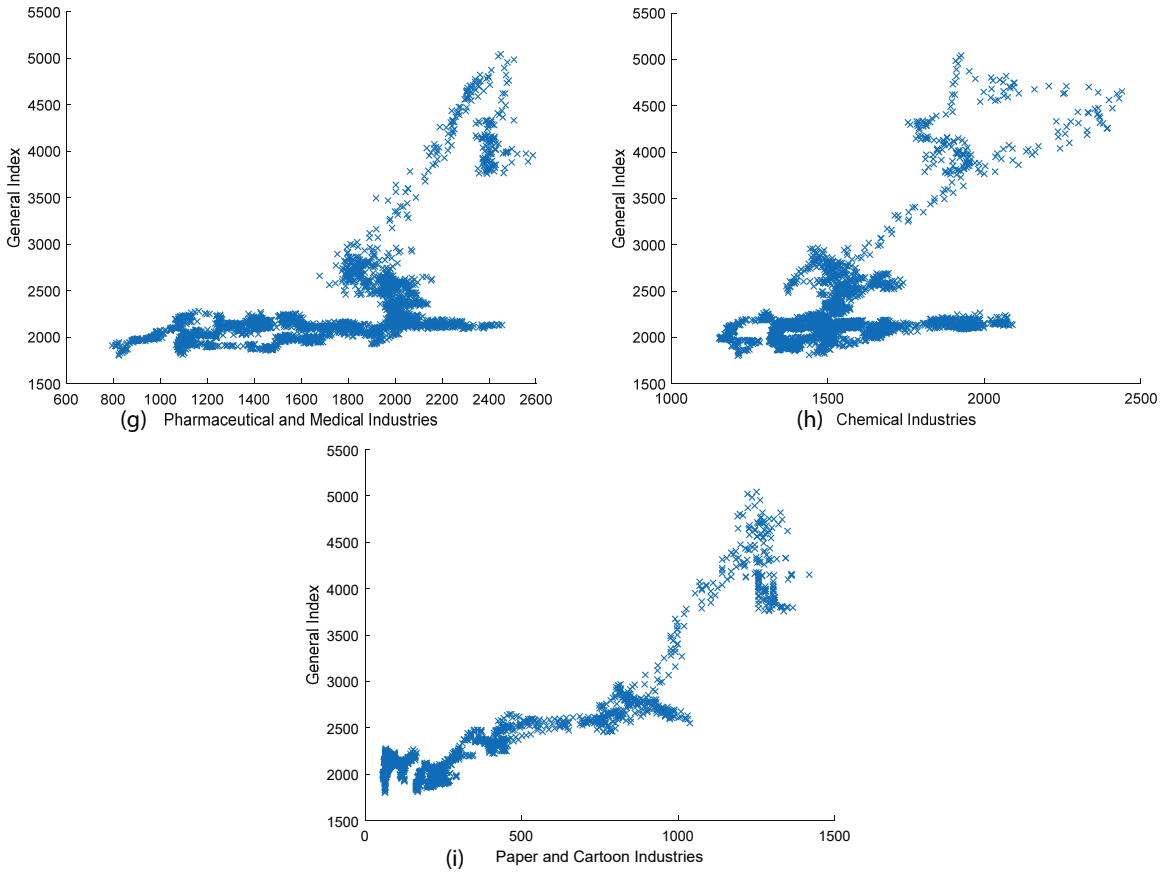
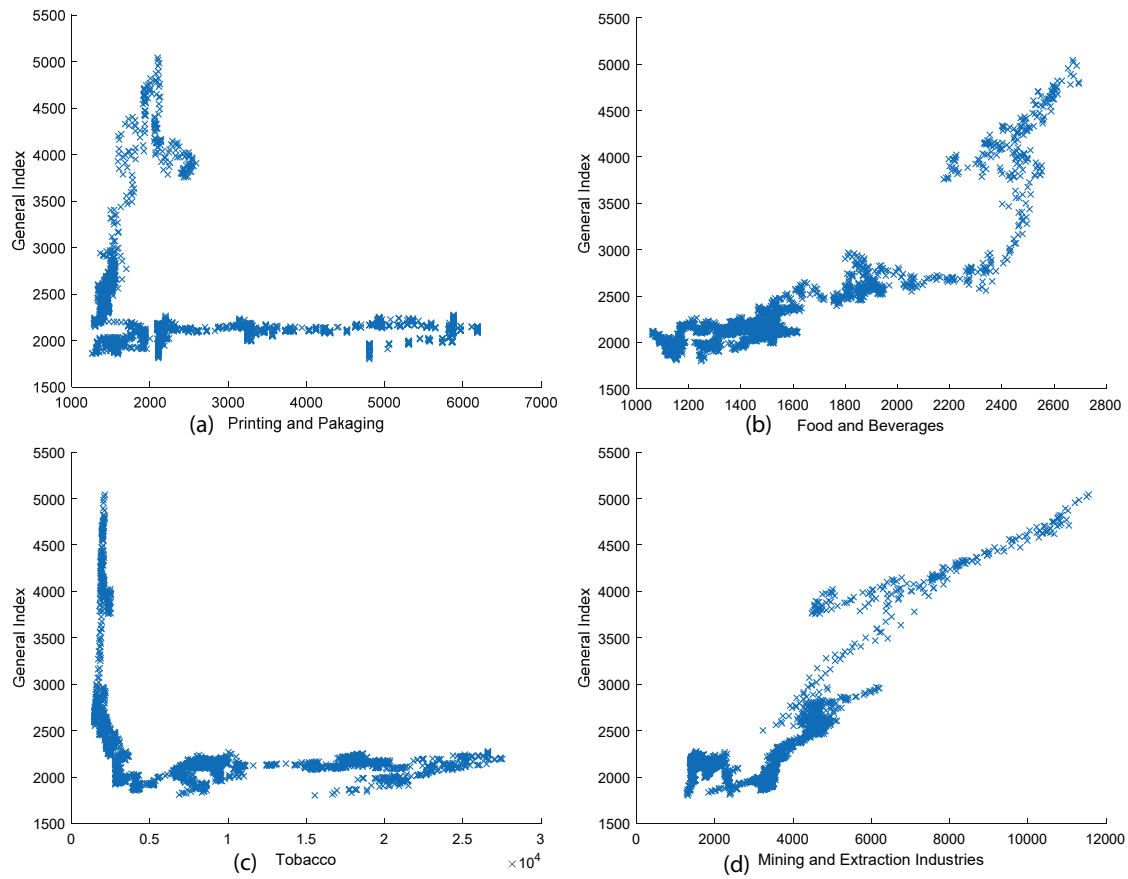


Fig. 3. *Relationship between the general index and different sector variables*



The end of Figure 4 on next page

The end of Figure 4

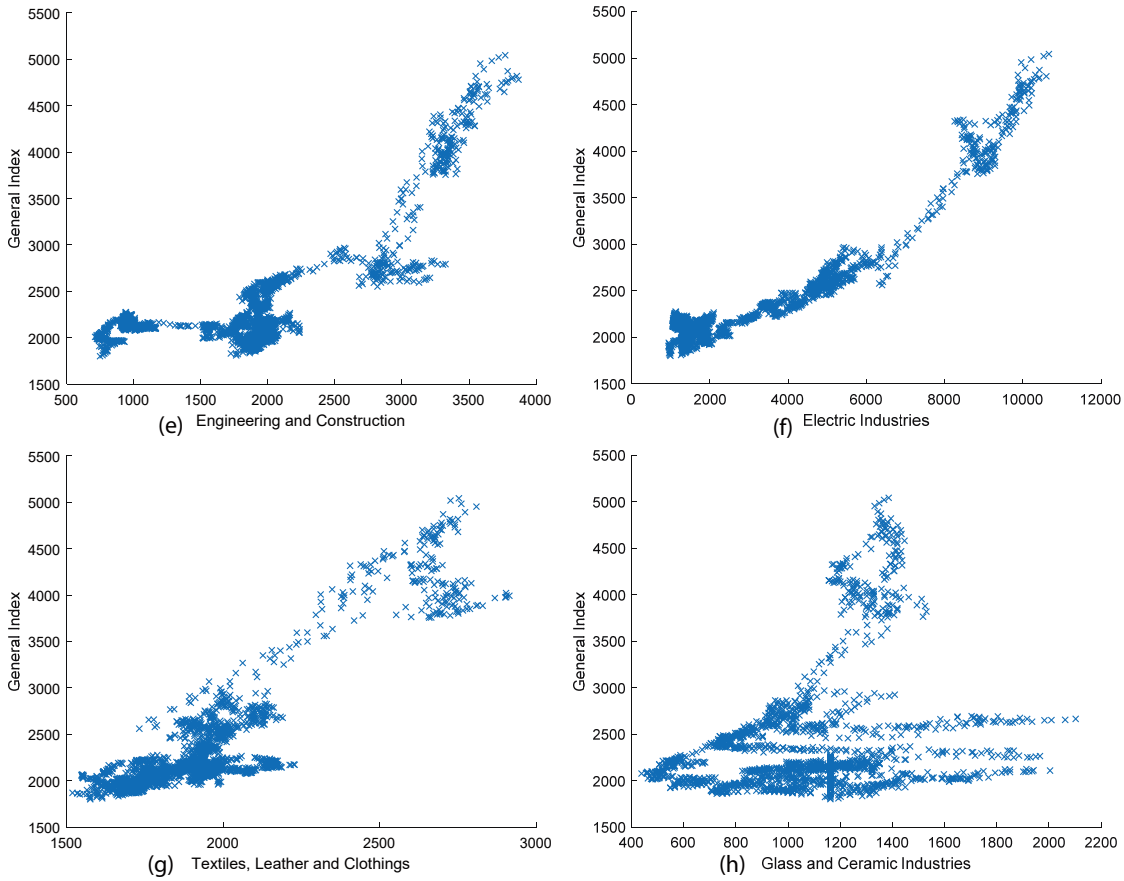


Fig. 4. *Relationship between general index and different sectors*

Table 5

Results of Model 2

Attribute	Rank	Attribute	Rank
Real Estate	0.9767	Banks	0.6936
Financial Services	0.9684	Technology and Communications	0.6117
Electric Industries	0.9392	Pharmaceutical and Medical Industries	0.5851
Insurance	0.9176	Chemical Industries	0.5465
Transportation	0.9059	Health Care Services	0.3621
Food and Beverages	0.8974	Glass and Ceramic Industries	0.2505
Paper and Cartoon Industries	0.8921	Day	-0.0113
Textiles, Leather and Clothing	0.8768	Month	-0.1245
Mining and Extraction Industries	0.8442	Educational Services	-0.1247
Commercial Services	0.8318	Printing and Packaging	-0.2263
Utilities and Energy	0.8231	Tobacco	-0.3739
Hotels and Tourism	0.8155	Year	-0.602
Media	0.7794	Banks	0.6936
Engineering and Construction	0.7387	Technology and Communications	0.6117

Utilities and Energy, and Hotels and Tourism are ranked more than 0.8. In addition, the two sectors, Media Engineering, and Construction, are ranked more than 0.7, while the rest of the attributes are less than 0.7.

Furthermore, the third model depended on ranker search and relief evaluator. Table 6 arranges the attributes from the most to least related. The top ten attributes are Day, Mining and Extraction Industries, Utilities and Energy, Banks, Health Care Services, Textiles, Leather and Clothing, Transportation, Chemical Industries, Financial Services, Technology and Communications. The rank values for the top ten variables are more than 0.007, whereas the rest of the attributes are less than 0.007, as shown in Table 6.

Table 6

Results of Model 3

Attribute	Rank	Attribute	Rank
Day	0.017	Engineering and Construction	0.007
Mining and Extraction Industries	0.016	Insurance	0.007
Utilities and Energy	0.016	Hotels and Tourism	0.006
Banks	0.016	Media	0.006
Health Care Services	0.011	Electric Industries	0.006
Textiles, Leather and Clothing	0.011	Month	0.005
Transportation	0.008	Paper and Cartoon Industries	0.005
Chemical Industries	0.008	Food and Beverages	0.005
Financial Services	0.008	Glass and Ceramic Industries	0.005
Technology and Communications	0.008	Pharmaceutical and Medical Industries	0.004
Commercial Services	0.007	Printing and Packaging	0.001
Real Estate	0.007	Year	0.000
Educational Services	0.007	Tobacco	-0.001

Moreover, the fourth model combines both ranker and classifier evaluators. The outcome of the test sorted the attributes from the strongest to the weakest in terms of their effect on the Amman Stock Exchange Index (ASEI), as shown in Table 7. The top ten attributes include Glass and Ceramic Industries, Real Estate, Educational Services, Hotels and Tourism, Transportation, Health Care Services, Financial Services, Textiles, Leather and clothing, Insurance, and Month.

Tables 5 to 7 showed the collected results from applying Models 1 to 4, respectively. Thus, only the highest frequency attributes will be considered to derive the best suitable combination of attributes from four models, as shown in Table 8.

Therefore, the criterion adopted for selecting the most effective attributes is mainly based on the number of times the attribute has occurred in each of the four models. The six attributes that ranked at least three times (i.e. three models out of four) were chosen to develop a linear regression model. The results encompass; Financial Services, Health Care Services, Mining and Extraction Industries, Real Estate, Transportation and Textiles, Leather and Clothing sectors; all six attributes are found to be the most affecting attributes on ASEI for the period of 2008–2018.

6.3 Building Linear Regression Model

A multiple regression model is built to test the reliability of the significant attributes found by applying multiple methods of Feature Selection. The OLS regression model results are shown in Table 9, the independent variables are Financial Services, Health Care Services, Mining and Extraction Industries, Real Estate, Transportation and Textiles, Leather and Clothing sectors. The R^2 value and the error value are 0.998 and 108, respectively.

Furthermore, Table 10 contains the ftest results indicating that the model is statistically significant. Besides that, Table 11 shows the output of ttest for each independent variable. Hence, the results show that all the independent variables are

Table 7

Results of Model 4

No.	Attribute	No.	Attribute
1	Glass and Ceramic Industries	14	Media
2	Real Estate	15	Utilities and Energy
3	Educational Services	16	Commercial Services
4	Hotels and Tourism	17	Mining and Extraction Industries
5	Transportation	18	Engineering and Construction
6	Health Care Services	19	Electric Industries
7	Financial Services	20	Tobacco
8	Textiles, Leather and Clothing	21	Food and Beverages
9	Insurance	22	Printing and Packaging
10	Month	23	Pharmaceutical and Medical Industries
11	Year	24	Chemical Industries
12	Banks	25	Paper and Cartoon Industries
13	Technology and Communications	26	Day

Table 8

The frequency of the most selected attributes from Models 1 to 4

Attribute	Frequency
Banks	1
Chemical Industries	1
Day	1
Educational Services	1
Electric Industries	1
Glass and Ceramic Industries	1
Hotels and Tourism	1
Month	1
Paper and Cartoon Industries	1
Technology and Communications	1
Utilities and Energy	1
Commercial Services	2
Food and Beverages	2
Insurance	2
Financial Services	3
Health Care Services	3
Mining and Extraction Industries	3
Real Estate	3
Transportation	3
Textiles, Leather and Clothing	4

Table 9

Model summary using OLS regression

R	R^2	Adjusted R^2	Error
0.999	0.998	0.998	108

Table 10

Analysis of variance test for linear regression

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Regression	15 705 086 528	3	5 235 028 843	139788	0
Residual	101 938 205	2722	37 450		
Total	15 807 024 734	2725			

Table 11

Coefficients of linear regression

Attribute	Beta	Sig.
Health Care Services	0.243	0.000
Mining and Extraction Industries	0.075	0.000
Textiles, Leather and Clothing	0.514	0.000
Real Estate	0.034	0.002
Financial Services	0.312	0.000
Transportation	-0.141	0.000

statistically significant, declaring a strong relationship between the dependent and independent variables. Thus, the results revealed that Financial Services, Health Care Services, Mining and Extraction Industries, Real Estate, Transportation and Textiles, Leather and Clothing sectors are the main sectors that affect the general index in the Jordanian stock market. Finally, the model is sufficient to be used for ASEI prediction.

7. Conclusion

This study is substantial because its main objective is to know the main features and sectors that strongly affect the general index of the Jordanian stock market ($ASEI_{100}$). In addition, the research focuses on how Jordan's stock market changed during the past ten years. The preliminary stage of analysis started by analysing and visualising each attribute against the general index ($ASEI_{100}$) and preparing correlation coefficients to understand the behaviour of $ASEI_{100}$ by using all sub-sectors in the stock market. The main findings after applying four feature selection methods for the period of 2008–2018 were that six sub-sectors are essential for $ASEI_{100}$ directly. Going through these sectors one by one, various justifications can be derived.

First, the six attributes share an essential common feature: they have a positive and fundamental effect on the economy. Moreover, financial services is an important sector in Jordan and the world due to its direct connection with financing and investment decisions for all types of firms. In addition, Jordan is a well-known exquisite country regarding health care services in the whole region, which in turn indicates that the performance of the health firms has a strong positive impact on the Jordanian economy, especially with the unstable political conditions in the neighbouring countries. Mining and extraction attribute is primary sectors related to natural resources. Jordan is known for being one of the top countries exporting resources (i.e., Phosphate, Potash, and chemical fertilisers). In addition, Real Estate, Textiles, Leather and Clothing, and Transportation sectors were affected directly by the increase in demand caused by the unstable political conditions of the region countries, especially in Iraq and Syria, which created the refugee crisis in Jordan.

The study is limited to the sub-sectors of the Jordanian stock market, so we recommend more features be taken into consideration regarding stock behaviour. This will promote the development of accurate and robust ASEI prediction models and give stakeholders the decision to fund and invest in ASEI[1].

[1] The data that support the findings of this study are openly available in Jordan Securities Commission at https://www.isc.gov.iq/, reference number data_2019.

References

Adebiyi, A. A., Adewumi, A. O. & Ayo, C. K. (2014). Comparison of ARIMA and artificial neural networks models for stock price prediction. *Journal of Applied Mathematics, 2014*.

Al-Najjar, D., Al-Najjar, H., Al-Rousan, N. & Assous, H. F. (2022). Developing Machine Learning Techniques to Investigate the Impact of Air Quality Indices on Tadawul Exchange Index. *Complexity, 2022*.

AL-Najjar, D., Al-Rousan, N. & AL-Najjar, H. (2022). Machine Learning to Develop Credit Card Customer Churn Prediction. *Journal of Theoretical and Applied Electronic Commerce Research, 17(4),* 1529-1542.

Al-Najjar, D., Assous, H. F., Al-Najjar, H. & Al-Rousan, N. (2022). Ramadan effect and indices movement estimation: a case study from eight Arab countries. *Journal of Islamic Marketing, (ahead-of-print)*.

Al-Najjar, H., Al-Rousan, N., Al-Najjar, D., Assous, H. F. & Al-Najjar, D. (2021). Impact of COVID-19 pandemic virus on G8 countries' financial indices based on artificial neural network. *Journal of Chinese Economic and Foreign Trade Studies, 14(1),* 89-103.

Assous, H. F., Al-Rousan, N., Al-Najjar, D. & Al-Najjar, H. (2020). Can international market indices estimate TASI's movements? The ARIMA model. *Journal of Open Innovation: Technology, Market, and Complexity, 6(2),* 27.

Bhuiyan, R. A., Rahman, M. P., Saiti, B. & Ghani, G. B. M. (2019). Does the Malaysian Sovereign Sukuk market offer portfolio diversification opportunities for global fixed-income investors? Evidence from wavelet coherence and multivariate-GARCH analyses. *The North American Journal of Economics and Finance, 47,* 675-687.

Cao, Q., Parry, M. & Leggio, K. B. (2011). The three-factor model and artificial neural networks: predicting stock price movement in China. *Annals of Operations Research, 185,* 25–44.

Chandar, S. K., Sumathi, M. & Sivanandam, S. N. (2016). Prediction of stock market price using hybrid of wavelet transform and artificial neural network. *Indian Journal of Science and Technology, 9(8),* 1-5.

Chauhan, N., Ravi, V. & Chandra, D. K. (2009). Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks. *Expert Systems with Applications, 36(4),* 7659-7665.

Chen, Y. & Hao, Y. (2018). Integrating principle component analysis and weighted support vector machine for stock trading signals prediction. *Neurocomputing, 321,* 381-402.

Chetty, N., Vaisla, K. S. & Sudarsan, S. D. (2015, December). Role of attributes selection in classification of Chronic Kidney Disease patients. In: *2015 International Conference on Computing, Communication and Security (ICCCS)* (pp. 1-6). IEEE.

Choudhry, T. & Osoble, B. N. (2015). Nonlinear interdependence between the US and emerging markets' industrial stock sectors. *International Journal of Finance & Economics, 20(1),* 61-79.

Chundakkadan, R. & Sasisdharan. S. (2019). Central bank's money market operations and daily stock returns. *International Journal of Finance & Economics, 26(1),* 136-152.

De Faria, E. L., Albuquerque, M. P., Gonzalez, J. L., Cavalcante, J. T. P. & Albuquerque, M. P. (2009). Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods. *Expert Systems with Applications, 36(10),* 12506-12509.

Gupta, R., Kanda, P., Tiwari, A. K. & Wohar, M. E. (2019). Time-varying predictability of oil market movements over a century of data: The role of US financial stress. *The North American Journal of Economics and Finance, 50,* 100994.

Hall, M. A. (1999). *Correlation-based feature selection for machine learning*. The University of Waikato, 198.

Hiransha, M., Gopalakrishnan, E. A., Menon, V. K. & Soman, K. P. (2018). NSE stock market prediction using deep-learning models. *Procedia computer science, 132,* 1351-1362.

Hu, H., Tang, L., Zhang, S. & Wang, H. (2018). Predicting the direction of stock markets using optimized neural networks with Google Trends. *Neurocomputing, 285,* 188-195.

Inthachot, M., Boonjing, V. & Intakosum, S. (2016). Artificial neural network and genetic algorithm hybrid intelligence for predicting Thai stock price index trend. *Computational intelligence and neuroscience, 2016*.

Khuat, T. T., Le, Q. C., Nguyen, B. L. & Le, M. H. (2016). Forecasting Stock Price using Wavelet Neural Network Optimized by Directed Artificial Bee Colony Algorithm. *Journal of Telecommunications and Information Technology, 2,* 43-52.

Kirikkaleli, D. (2018). The effect of domestic and foreign risks on an emerging stock market: A time series analysis. *The North American Journal of Economics and Finance, 51,* 100876.

Kononenko, I. (1994). Estimating attributes: analysis and extensions of RELIEF. In: *European conference on machine learning* (pp. 171-182). Springer, Berlin, Heidelberg.

Lo, A. W. & Mackinlay, C. (1988). Stock Market Prices do not follow random walk: evidence from a simple specification test. *The review of financial studies, 1,* 41-66.

Moghaddam, A. H., Moghaddam, M. H. & Esfandyari, M. (2016). Stock market index prediction using artificial neural network. *Journal of Economics, Finance and Administrative Science, 21(41),* 89-93.

Nahil, A. & Lyhyaoui, A. (2018). Short-term stock price forecasting using kernel principal component analysis and support vector machines: the case of Casablanca stock exchange. *Procedia Computer Science, 127,* 161-169.

Onik, A. R., Haq, N. F., Alam, L. & Mamun, T. I. (2015). An analytical comparison on filter feature extraction method in data mining using J48 classifier. *International Journal of Computer Applications, 124(13)*.

Orimoloye, L. O., Sung, M. C., Ma, T. & Johnson, J. E. (2020). Comparing the effectiveness of deep feedforward neural networks and shallow architectures for predicting stock price indices. *Expert Systems with Applications, 139,* 112828.

Pan, Y., Xiao, Z., Wang, X. & Yang, D. (2017). A multiple support vector machine approach to stock index forecasting with mixed frequency sampling. *Knowledge-Based Systems, 122,* 90-102.

Qiu, M. & Song, Y. (2016). Predicting the direction of stock market index movement using an optimized artificial neural network model. *PloS One, 11(5),* e0155133.

Rather, A. M., Agarwal, A. & Sastry, V. N. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications, 42(6),* 3234-3241.

Ruiz, R., Riquelme, J. C. & Aguilar-Ruiz, J. S. (2005, June). Heuristic search over a ranking for feature selection. In: *Computational Intelligence and Bioinspired Systems. 8th Work-Conference on Artificial Neural Networks* (pp. 742-749). Springer, Berlin, Heidelberg.

Syed Nor, S. H., Ismail, S. & Yap, B. W. (2019). Personal bankruptcy prediction using decision tree model. *Journal of Economics, Finance and Administrative Science, 24(47),* 157-170.

Tang, H., Dong, P. & Shi, Y. (2019). A new approach of integrating piecewise linear representation and weighted support vector machine for forecasting stock turning points. *Applied Soft Computing, 78,* 685-696.

Umer, U. M., Sevil, T. & Sevil, G. (2019). Forecasting performance of smooth transition autoregressive (STAR) model on travel and leisure stock index. *The Journal of Finance and Data Science, 4(2),* 90-100.

Wang, J., Wang, J., Fang, W. & Niu, H. (2016). Financial time series prediction using Elman recurrent random neural networks. *Computational intelligence and neuroscience, 2016.*

Zahedi, J. & Rounaghi, M. M. (2015). Application of artificial neural network models and principal component analysis method in predicting stock prices on Tehran Stock Exchange. *Physica A: Statistical Mechanics and its Applications, 438,* 178-187.

Zhou, F., Zhou, H. M., Yang, Z. & Yang, L. (2019). EMD2FNN: A strategy combining empirical mode decomposition and factorization machine based neural network for stock market trend prediction. *Expert Systems with Applications, 115,* 136-151.

About the authors

Dana Al-Najjar — Dr., Associate Professor of the Department of Finance and Banking Sciences, Faculty of Business, Applied Science Private University; https://orcid.org/0000-0001-7292-1536 (Amman, 11931, Jordan; e-mail: d_alnajjar@asu.edu.jo).

Hazem Al-Najjar — Dr., Assistant Professor of the Department of Computer Engineering, Faculty of Engineering and Architecture, Istanbul Gelisim University; https://orcid.org/0000-0002-6143-2734 (Istanbul, 34310, Turkey; e-mail: hazem_najjar@yahoo.com).

Nadia Al-Rousan — Dr., Assistant Professor of MIS Department, Faculty of Business, Sohar University; https://orcid.org/0000-0001-8451-898X (Sohar, 311, Oman; e-mail: nadia.rousan@yahoo.com).

Информация об авторах

Аль-Наджар Дана — доктор наук, доцент кафедры финансов и банковских наук, факультет бизнеса, Частный университет прикладных наук; https://orcid.org/0000-0001-7292-1536 (Иордания, 11931, г. Амман; e-mail: d_alnajjar@asu.edu.jo).

Аль-Наджар Хазем — доктор наук, доцент кафедры компьютерной инженерии, факультет архитектуры и строительства, Стамбульский университет Гелишим; https://orcid.org/0000-0002-6143-2734 (Турция, 34310, г. Стамбул; e-mail: hazem_najjar@yahoo.com).

Аль-Роусан Надя — доктор, доцент кафедры ИСУ, факультет бизнеса, Сохарский университета; https://orcid.org/0000-0001-8451-898X (Оман, 311, г. Сохар; e-mail: nadia.rousan@yahoo.com).

Дата поступления рукописи: 10.08.2020. Received: 10 Aug 2020.
Прошла рецензирование: 20.10.2020. Reviewed: 20 Oct 2020.
Принято решение о публикации: 15.09.2022. Accepted: 15 Sep 2022.