Dana Al-Najjar [a] (iD)*, Hazem Al-Najjar* [b] (iD) ✉*, Nadia Al-Rousan* [c] (iD)
[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

---

Д. Аль-Наджар ᵃ⁾ (iD) *, Н. Аль-Наджар* ᵇ⁾ (iD) ✉*, Н. Аль-Роусан* ᵛ⁾ (iD)

ᵃ⁾ Частный университет прикладных наук, г. Амман, Иордания
ᵇ⁾ Стамбульский университет Гелишим, г. Стамбул, Турция
ᵛ⁾ 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

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 gen-

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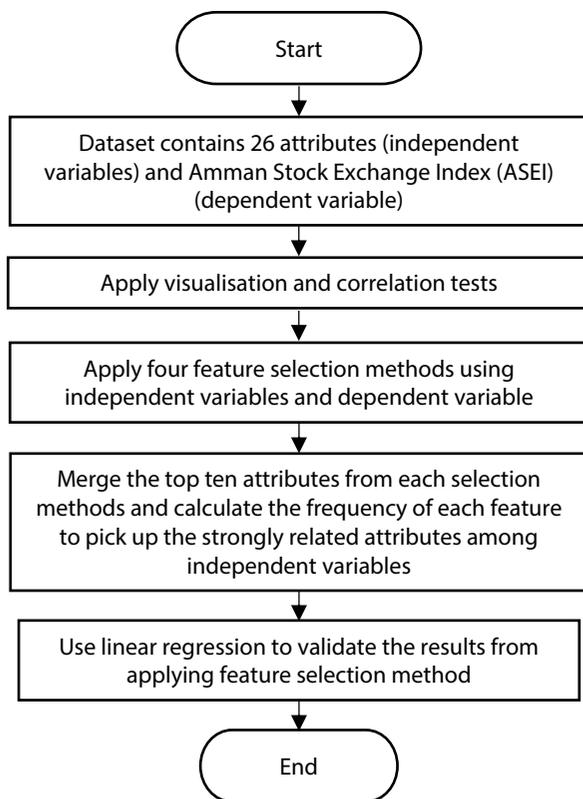
2. The free-float market capitalisation of each firm is calculated by multiplying the market capitalisation with the free-float factor that lies between 0.05 and 1.00. The equation to calculate the free-float factor is the shares owned by the board of directors and investors who own more than 5 % of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

```
        ┌─────────────────────┐
        │        Start        │
        └─────────────────────┘
                  │
    ┌─────────────────────────────────┐
    │ Dataset contains 26 attributes   │
    │ (independent variables) and      │
    │ Amman Stock Exchange Index (ASEI)│
    │ (dependent variable)             │
    └─────────────────────────────────┘
                  │
    ┌─────────────────────────────────┐
    │ Apply visualisation and          │
    │ correlation tests                │
    └─────────────────────────────────┘
                  │
    ┌─────────────────────────────────┐
    │ Apply four feature selection     │
    │ methods using independent        │
    │ variables and dependent variable │
    └─────────────────────────────────┘
                  │
    ┌─────────────────────────────────┐
    │ Merge the top ten attributes     │
    │ from each selection methods and  │
    │ calculate the frequency of each  │
    │ feature to pick up the strongly  │
    │ related attributes among         │
    │ independent variables            │
    └─────────────────────────────────┘
                  │
    ┌─────────────────────────────────┐
    │ Use linear regression to         │
    │ validate the results from        │
    │ applying feature selection method│
    └─────────────────────────────────┘
                  │
        ┌─────────────────────┐
        │         End         │
        └─────────────────────┘
```

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 be-

haviour 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