Alexandria Engineering Journal (2019) 58, 499–506

HOSTED BY

ELSEVIER

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

ORIGINAL ARTICLE

Estimation and prediction of construction cost index using neural networks, time series, and regression

Yasser Elfahham

Structural Engineering Dep., Alexandria University, Egypt

Received 1 August 2018; revised 30 November 2018; accepted 8 May 2019
Available online 23 May 2019

KEYWORDS

Construction cost index;
Inflation;
Estimating;
Forecasting;
Neural networks;
Regression;
Time series

Abstract Estimating construction costs and predicting price escalation are major steps for project owners, estimators, and contractors. The construction costs are always subject to fluctuations that trend toward increasing over the long term, which make the pricing process challenging job. The construction cost index (CCI) has been widely used to forecast project costs. The problem is that no agency provides estimation for that important index in Egypt. In this paper, a formula is driven for calculating the Construction Cost Index for concrete structures based on past records of key construction costs. Neural Networks, Linear Regression, and Autoregressive Time Series are then utilized to forecast the CCI. The main contribution of this study is providing construction stakeholders with a reliable tool for expecting prices of coming projects, especially with the existing Rates of Inflation.

1. Introduction

The Construction Cost Index (CCI) is a weighted aggregate index of the prices of constant quantities of materials [15]. This index provides changes occurring in costs on both short- and long terms in an attempt to get more accurate bids [16]. Owners need this index to get the probable project costs, while contractors use it in tender phase to submit their financial offers.

There are different factors should be examined when constructing an index. These factors include; purpose of the index, selection of cost elements, choice of weights of each element, and choice of base year [18]. Several cost indices were developed and utilized for this purpose, among them, Engineering News Record (ENR) in North America which is issued at weekly base. ENR is a weighted aggregate index of the prices of constant quantities of structural steel, Portland cement, Lumbers and Common Labor [10].

In Egypt, "Central Agency for Public Mobilization and Statistics" [6] is the agency responsible for statistics and data collection, it also publishes periodical reports that provide the costs of construction materials and other indicators. However, these numbers are the costs of individual items and do not reflect how much the construction costs are decreasing or increasing [12]. That shows the need to generate an index that integrates the basic costs of major construction component.

The objective of this paper is to estimate the value Construction Cost Index based on historical data of key cost items

E-mail address: yelfahham@yahoo.com

Peer review under responsibility of Faculty of Engineering, Alexandria University.

https://doi.org/10.1016/j.aej.2019.05.002
1110-0168

and provide a reliable forecasting tool for predicting CCI at any year in the future. The motive for this study is that there is no statistical agency estimates or publishes the CCI in Egypt like that in developed countries. The study is limited to Reinforced concrete structures. The study starts by generating a calculation formula for CCI, and then, forecasting the future values of CCI based on historical data of cost items. This paper is further comprised of the following sections: literature review, research methodology, calculating construction cost index, predicting construction cost index, evaluation of forecasting accuracy and conclusion and recommendations.

2. Literature review

The construction costs are dynamic. Prices of materials, human resources, and other costs fluctuate continuously. This economic uncertainty could have a major impact on the business especially on long term and mega projects. To limit potential financial risks, a clear consideration is required. Construction cost indices are used to predict expected costs.

Many forecasting techniques have been developed in recent years to handle the complexity of prediction problems. Three basic approaches are existed for construction cost prediction, (1) the traditional methods, (2) the modern methods and (3) the Quantitative methods. The traditional method must specify the relationships between the predicted or dependent variable, and the independent variables [13]. Example of traditional method is Regression method. Previous researchers Cheung & Skitmore [8], Trost & Oberlender [17], Chen [7], Hammad et al. [11] have predicted the future CCI based on the traditional method. The main disadvantage of the traditional method is that the all explanatory variables must be identified to forecast the dependent variable. This approach was not popular in the construction industry as too many independent variables can affect construction cost [1].

In contrast, modern methods represented in time series analysis, is a series of data points listed at equally spaced points in time order [9]. Time series methods attempt to predict the future data values of a series based on analyzing previous data values using the internal statistical between data [4]. Chijoo Lee et al. [5] proposed a multivariate time series analysis to predict the raw material prices with a focus on iron ore as it presents the main contributor to the prices of steel products. They also examined the effect of accurate predictions of project profitability. The results were more accurate by 2.3 times compared to past records.

The Quantitative methods like Artificial Neural Networks (ANN) is one of the Artificial intelligence computing systems inspired by the human brain as ANN simulates the learning ability of brains. Neural networks were applied to forecast the escalation occurred in the costs of highway projects with reasonable accuracy [19]. Trefor [16] predicted fluctuations in CCI by applying three different approaches: ANN, exponential smoothing, and multiple regressions, the forecasted results of ANN were less accurate than other methods. Kim [14] showed that ANN is most beneficial for longer-term forecasts than other statistical methods due to the lack of historical data. Some studies tried to combine more prediction tools in one model. Byung-Cheol and Young [3] presented a forecast combination model to detect early potential cost overrun of project. The model identifies the best forecasting model commonly used and visualizes the likely errors. The results of applying the model on real project revealed high accuracy while mitigating the risk of large errors. Table 1 provides the common prediction methods previously discussed with their main pros and cons.

3. Methodology

In this study, a formula for calculating the Construction Cost Index (CCI) is derived for reinforced concrete structures with the focus on key used materials. The prices of the used materials are obtained from historical records. Next, a prediction for the generated CCI is performed using three different approaches. The prediction results of the three methods are then analyzed to evaluate the performance of them and find out the best method that provides the higher accuracy. Fig. 1 provides a brief flowchart of the research Objective and Methodology.

4. Data collection

The calculation formula of CCI is based on the “Engineering News Record” (ENR) index as it contains the basic cost items and construction materials commonly used in North America. However, some modifications are done to fit the construction nature and materials used in Egypt. ENR index is calculated using the prices of Structural Steel, Portland cement and Lumber, because these are the most materials affecting construction costs in America. On the other hand, Lumber is rarely used as a construction material, most Buildings are constructed from Reinforced Concrete skeleton and bricks, therefore, other

Table 1 Forecasting techniques [1].

Method type	Forecasting method	Technique	Advantages	Disadvantages
Econometric methods	Traditional	Multiple regression analysis	• Easy implementation by practitioners • Good results can be gained with fairly small data set	• All the independent variables to be identified
	Modern	Times series methods	• Typically one variable is needed in order to forecast	• difficult • Historical data set is needed
Non econometric methods	Quantitative	Neural networks	• Less statistical training is required • Ability to address complex non-linear relationships	• Its black-box nature

materials that are extensively used in construction have to be included. The proposed CCI will use the prices of structural steel, Portland cement, bricks, sand and gravel as key cost items as it will be discussed in the following sections.

The historical prices of the selected key cost items discussed previously were collected from ''the Central Agency for Public Mobilization and Statistics [6]''. The available data about these materials provides the total locally produced quantities and corresponding prices in Egyptian pounds (LE). The average unit price of each material is then calculated through dividing the produced quantity by their corresponding price. The collected data is 16 years (year 2002–2018) on a yearly basis to provide clear picture about economic changes each year. It is well known that the Egyptian economy is continuously changing and is not stable, especially, in the last decade. Fig. 2 presents the total prices of the materials used in estimating the proposed CCI during the study period (2002–2018) where the X - axis represents the study year, while the Y - axis represents the yearly prices of the total quantity produced at that year. Fig. 3 provides the unit price of the four materials. The unit price is calculated by dividing the total yearly price of a material over its annual production quantity. It can be concluded from Fig. 3 that Cement and Steel were the most materials that had rapid rates of price increase while Sand and Gravel were the least especially during the years 2007, 2008, and after 2014.

Fig. 3 also shows that the material prices often fluctuate significantly which in turn affects the construction cost.

5. Calculating construction cost index (CCI)

This section derives a calculation formula for the proposed CCI. The formula should include the key cost items discussed previously with the suitable weight for each cost item. First of all, a base year within the study period should be identified first. Selecting a base year is important as it affects the accuracy of generated formula. The prices of used materials are

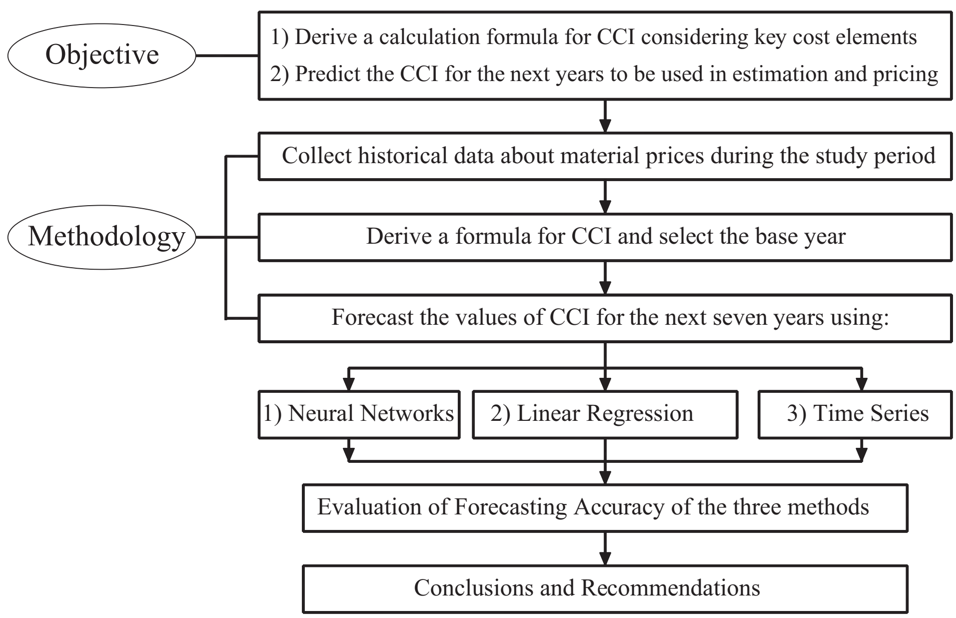

Fig. 1 Research objective and methodology.

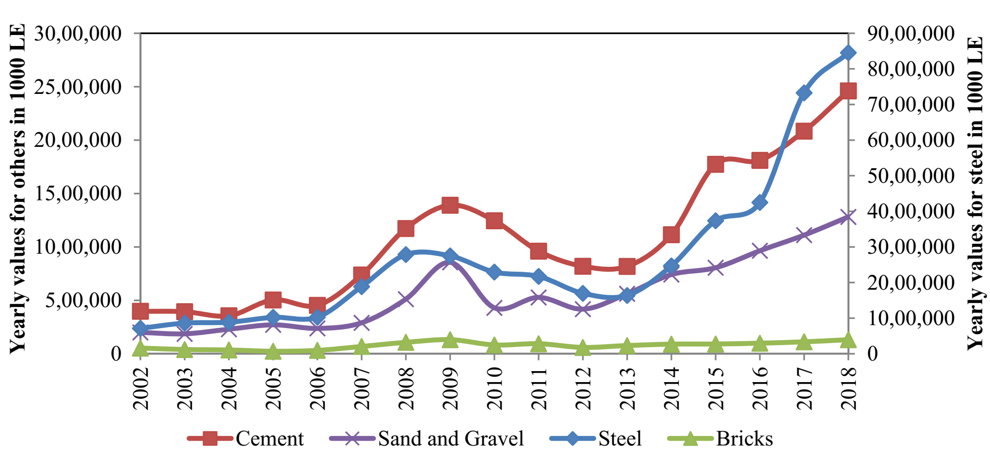

Fig. 2 Total prices (in 1000 LE) for the four materials used in the construction cost index.

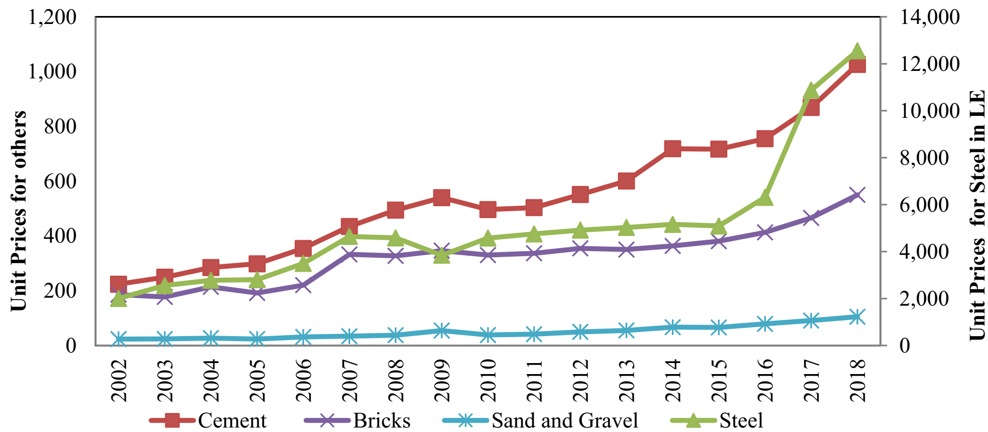

Fig. 3 Unit prices for the four material components used in the construction cost index.

analyzed for that purpose; a year with least fluctuations in prices is preferred which shows less inflation rates and thus more economic stability within the study period. The chart in Fig. 3 is analyzed by calculating the average price slope between each two successive points to recognize the year with the most stability in prices. Years with negative slopes are excluded from the analysis. The result of the analysis showed that year 2010 was the most stable year and therefore is was selected as base year.

Afterwards, the weights for construction cost index were calculated by determining the quantities of the four material components based on their production values and unit prices during the base year. The formulas below show in details how the weights of Construction Cost Index (CCI) are calculated for the base year:

- Eq. (1): Bricks weight in 1000 bricks in construction cost index (BW):

$$BW = \frac{BP_b \times 100/(BP_b + SP_b + CP_b + GP_b)}{BU_b} \tag{1}$$

- Eq. (2): Steel weight in ton in construction cost index (SW):

$$SW = \frac{SP_b \times 100/(BP_b + SP_b + CP_b + GP_b)}{SU_b} \tag{2}$$

- Eq. (3): Cement weight in ton in construction cost index (CW):

$$CW = \frac{CP_b \times 100/(BP_b + SP_b + CP_b + GP_b)}{CU_b} \tag{3}$$

- Eq. (4): Sand and gravel weight in m^3 in construction cost index (GW):

$$GW = \frac{GP_b \times 100/(BP_b + SP_u + CP_b + GP_b)}{GU_b} \tag{4}$$

where

BP_b: Bricks production at the base year in LE.
SP_b: Steel production at the base year in LE.
CP_b: Cement production at the base year in LE.
GP_b: Sand and gravel production at the base year in LE.
BU_b: Bricks unit price at the base year in LE/1000 bricks.
SU_b: Steel unit price at the base year in LE/ton.
CU_b: Cement unit price at the base year in LE/ton.
GU_b: Sand and gravel unit price at the base year in LE/m^3.

The weights of each material used in calculating the CCI is obtained from Eqs. (1)–(4). The values of these weights for the base year are obtained as follows:

$BW = 0.0178$ in 1,000 bricks, $SW = 0.0119$ in tons, $CW = 0.0595$ in tons, and $GW = 0.2661$ in m^3.

The material components weights of the Construction cost index have to be converted into percentages; this is accomplished by multiplying each material's weight by its unit price at the base year. Fig. 4 shows these percentages; it can be observed from Fig. 4 that Steel has the highest weight in calculating the CCI with a percentage of 54% followed by cement, sand and gravel, while the smallest weight is for bricks. Therefore, the Steel and Cement have the major effect on the concrete structure with a percentage of 84%.

After knowing the weights of the materials involved in calculating the CCI, the Index can now be calculated for the remaining years of the study period from 2002 to 2018 by multiplying each material's weight by its unit price using Eq. (5) shown below. Table 2 shows the values of the CCI calculated from Eq. (5) based on the data of CAMPAS.

Eq. (5): Construction cost index (CCI) for the year under consideration

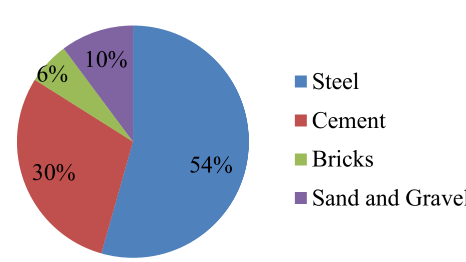

Fig. 4 Percentages of the four material components used in the construction cost index for the base year (2010).

Table 2 Values of CCI for the study period (2002–2018).

Construction cost index (CCI)	46.67	54.74	60.95	60.75	74.91	96.02	99.77	98.52	100	103.45	110.81	116.37	128.32	127.53	148.29	213.56	248.02
Year	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018

$$CCI = 0.0178 \times BU_i + 0.0119 \times SU_i + 0.0595 \times CU_i + 0.2661 \times GU_i \quad (5)$$

where

BU_i: Brick unit price for year i.
SU_i: Steel unit price for year i.
CU_i: Cement unit price for year i.
GU_i: Sand and gravel unit price for year i.

6. Predicting construction cost index

The construction cost index (CCI) is calculated in the previous sections, this section shows the prediction procedures of the CCI using three different approaches. The accurate prediction of the CCI is necessary for better estimation of project budgets, especially with those project that take many years where construction costs are continuously increased over time. In addition, bidders also can benefit from the predicted CCI by providing more reasonable financial offers. In this section, three different forecasting methods, Neural Network, Regression, and Autoregressive Time Series methods were applied to forecast construction cost index then their results were compared. The forecasted period is selected to be seven years after the current year; therefore the CCI can be generated until 2025.

6.1. Prediction using neural networks

Artificial Neural Networks (ANN) is extensively used to predict the future of any data type given the historical data. ANN uses algorithms and computing processes similar to the structure and functions neural systems of the human brain. ANN as nonlinear statistical data modeling tools learns from input data to predict the outputs where the relationships between inputs and outputs are modeled or patterns are found. In this study, NeuroXL Clusterizer software is used where it uses self-organizing neural networks.

In order to perform the neural networks calculations, two stages have to be done; the first stage is learning the algorithm, while the second stage is testing it. For that purpose; the data of the CCI form 2002 to 2018 is divided into two sets, the first set contains 95% of the data assigned for training; while the remaining 5% assigned for testing.

The results of the testing stage showed that neural networks can provide acceptable performance and give predicted Construction cost index with small average errors. Fig. 5 gives the predicted profile of the CCI using ANN. It was noticed that over the study period, the construction cost index had almost stable trend except for the last three years (2016, 2017, and 2018) that shows unusual increase or jump due to the economic decisions taken through this period by the Egyptian government based on the guidance of the World Bank. These economic decisions led to exceptional increase in prices and inflation rates. The broken lines from 2002 to 2018 provide the calculated CCI, while the dotted curve after 2018 provides the predicted CCI.

6.2. Prediction using regression method

Linear Regression Method is a simple approach commonly used in modeling the relationship between scalar variables and by knowing the trend of the data, prediction could be known. The Construction Cost Index was forecasted by using the least square linear regression method. That is accomplished through calculating the value of R^2. For the given data, the value of R^2 is 0.81. Fig. 6 provides the prediction of CCI using Linear Regression Method. As it can be noted from Fig. 6, the Linear Regression Method produces linear trend of the predicted CCI and it is not logic as it is represents a theoretical case. It can be noted also that the predicted line declines after year 2018, which is not expected, and takes linear trend after

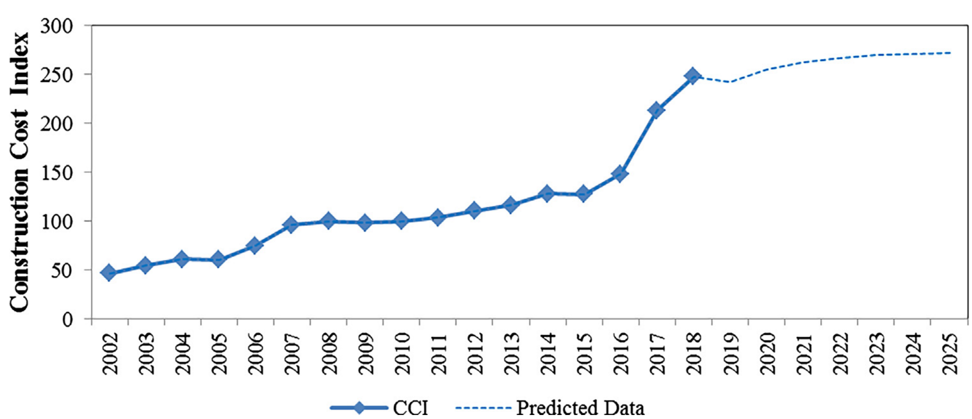

Fig. 5 Construction cost index prediction using neural networks.

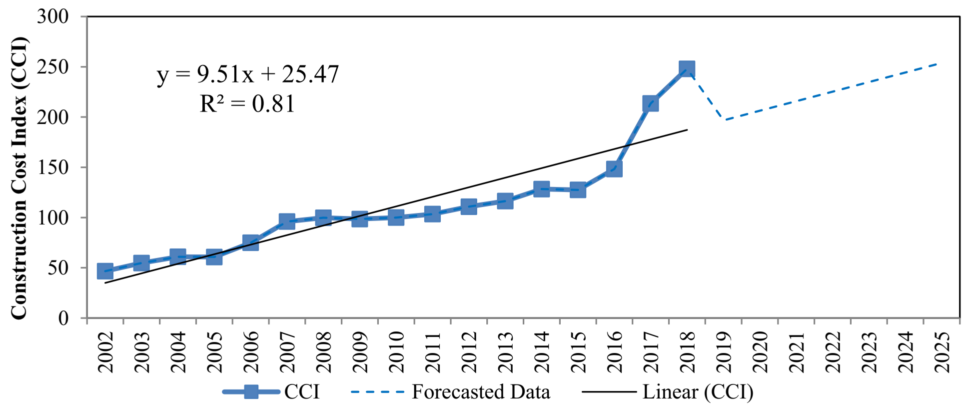

Fig. 6 Construction cost index prediction using regression method.

that date. Therefore, it can be said that Linear Regression Method provides only the approximate trend of the predicted variable. Additional analysis of prediction performance of this method will be discussed later.

6.3. Prediction using autoregressive time series

Time-series methods make forecasts based on historical patterns in the data. Time-series methods use time as independent variable to produce the other variables. Autoregressive Time Series is a common approach for modeling univariate time series. The Construction Cost Index is forecasted by using the autoregressive (AR) model.

The AR (p) model is defined as:

$$X_t = c + \sum_{i=1}^{p} \varnothing_i X_{t-i} + \varepsilon_t$$

where p is the order of the model; $\varnothing_1, \cdots\cdots\cdots, \varnothing_p$ are coefficients of the model; c is a constant; and $\varepsilon_t$is white noise.

MATLAB R2015a is used as tool to develop the proposed Autoregressive Time Series model to provide prediction of the CCI index from 2019 to 2025. The output data was plotted. It was observed that the CCI reached about 336.03% by 2025, as shown in Fig. 7 which provides predicted CCI using Autoregressive Time Series method.

7. Evaluation of forecasting accuracy

In this section, the results of the prediction methods are evaluated to measure the accurecy of them and find out which method provides higher performance. For that purpose, the mean squared errors and the mean absolute errors approaches are used as recommended by Jiang & Sungwoo [13]. It worth noting that smaller values of mean squared errors and the mean absolute errors indicate better forecasting ability. Given the data set of the construction cost index for the period of 2002–2018, and the forecasting using the proposed three methods; where the results can be compared as shown Table 3 and Fig. 8.

Table 3 Values of predicted CCI for using three prediction methods.

Prediction year	Predicted Construction Cost Index (CCI)		
	Neural networks	Linear regression	Autoreg. time series
2019	242.79	197.00	272.17
2020	254.81	206.00	289.65
2021	258.70	210.00	302.81
2022	262.52	215.76	313.19
2023	267.10	225.28	321.79
2024	269.70	234.79	329.26
2025	271.15	244.31	336.03

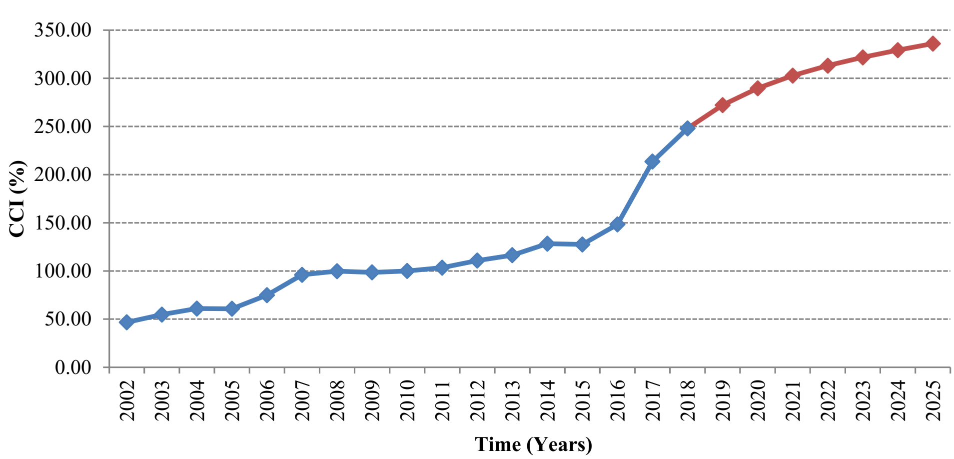

Fig. 7 Prediction of CCI using Autoregressive Time Series method.

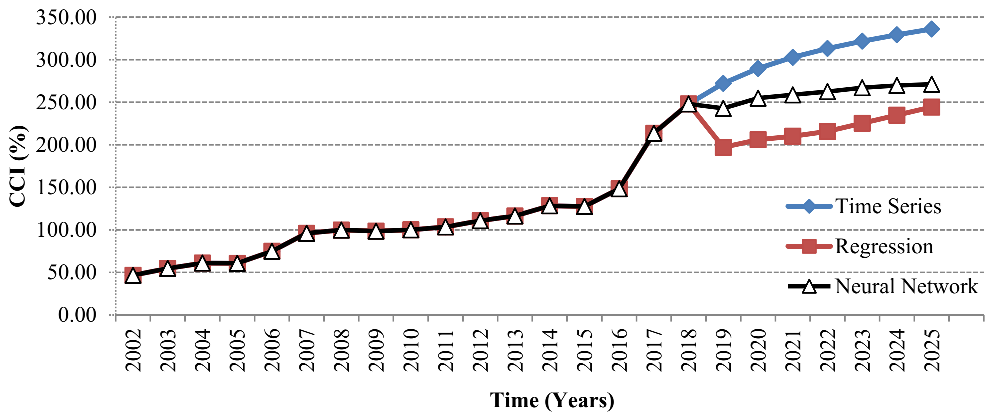

Fig. 8 Comparison of predicted CCI using three different methods.

It can be noted that the prediction results are reasonable to some extent, but the autoregressive time series was the best. The average of absolute errors is calculated and they are: 8.3 for Neural Network; 17.5 for Regression method, which was the less accurate method; whereas it is 3.5 for the time series method and thus is the best.

8. Conclusions and recommendations

Construction projects require high capital investment and may take many years to be completed. That makes them subject to prices escalation. Construction cost index (CCI) is an index that provides the fluctuations in construction costs. This is beneficial in evaluating, pricing and bidding construction projects, especially with the high rates of Inflation occurs in Egypt. In this paper, the CCI is calculated on the bases of Engineering News Record Index in America with the difference of changing the cost items used in calculation formula to suit the construction materials used in Egypt, namely; the study is limited to concrete structures only. That data was collected from the "Central Agency for Public Mobilization and Statistics" [6], which represents the leading statistical agency. The analysis of the collected data revealed that the key cost items concrete structures are Steel, Cement, Sand and Gravel, and Bricks. The weights of these material components in the formula were found to be 54%, 30%, 10%, and 6% respectively. The CCI was calculated for the study period from 2002 to 2018 considering that 2010 is the base year as shown in Table 2.

Neural network, Linear Regression and autoregressive time series methods were used to predict the construction cost index until 2025 as shown in Table 3 and Fig. 8. The performance of the three forecasting methods was then evaluated using the mean squared errors and the mean absolute errors approaches. It was found that the prediction using Autoregressive Time Series was found to be the most accurate method, as it produced the least average of absolute errors of 3.5, while Regression method was less effective with average of absolute errors of 17.5, besides, Linear Regression Method provides linear prediction which is not practical.

The main contribution of this study is presenting the CCI which could be used to predict the escalation occurred in construction costs of upcoming projects with acceptable level of accuracy. That could enhance the cost estimation process for project estimators and mitigate the risk associated with the inflation rates, which in turn hinders projects from progressing and create more claims between project stakeholders. Recommendations for future works includes using additional cost items like construction Labor and Equipment, and using other forecasting techniques.

References

[1] J. Alireza, N. Isabelina, Volatility forecast of construction cost index using general autoregressive conditional heteroskedastic method, J. Constr. Eng. Manage. ASCE 142 (1) (2016) 1–12.
[3] K. Byung-Cheol, H.K. Young, Improving the accuracy and operational predictability of project cost forecasts: an adaptive combination approach, Prod. Plann. Control 29 (9) (2018) 743–760.
[4] A. Baabak, Time series analysis of ENR construction cost index, J. Constr. Eng. Mange. ASCE 136 (11) (2010) 1227–1237.
[5] Chijoo Lee, W. Jongsung, Eul-Bum Lee, Method for predicting raw material prices for product production over long periods, J. Constr. Eng. Manage. ASEC 145 (2019) 1.
[6] CAPMAS, Economic Statistics for Arab Republic of Egypt – Building and Construction Statistics for the Public/Publics Work Sectors, Public Information Center Printing Office, Egypt, 2018.
[7] H. Chen, Developing cost response models for company-level cost flow forecasting of project-based corporations, J. Manage. Eng. 23 (4) (2007) 171–180.
[8] S. Cheung, R. Skitmore, Prediction of tender price index directional changes, Constr. Manage. Econ. 18 (7) (2000) 843–852.
[9] W. Enders, Applied Econometric Time Series, Wiley, New York, 2008.
[10] T. Gorgan, Methodology: inflation swing back up. Engineering News Record, 27 March, 2000, p.110.
[11] A. Hammad, S. Ali, G. Sweis, Statstical analysis on the cost and duration of public building projects, J. Manage. Eng. 26 (2) (2010) 105–113.
[12] A. Hassanein, Khalil, Building Egypt 1 – a general indicator cost index for the Egyptian construction industry, J. Eng., Constr. Archit. Manage. 13 (5) (2006) 463–480.
[13] W.X. Jiang, M. Sungwoo, Stochastic forecast of construction cost index using a cointegrated vector autoregression model, J. Manage. Eng. ASCE 29 (1) (2013) 10–18.
[14] G. Kim, Comparison of construction cost estimating models based on regression analysis, neural networks and case-based reasoning, J. Build. Environ. 39 (10) (2004) 1235–1242.

[15] C. Minh-Tu, C. Min-Yuan, Yu-Wei, Hybrid computational model for forecasting Taiwan construction cost index. J. Constr. Eng. Manage. ASCE, 141 (2015) (4), pp.1-11.
[16] W.P. Trefor, Predicting changes in predicting changes in construction cost indexes using neural networks, J. Constr. Eng. Manage. ASCE 120 (2) (1994) 306–320.
[17] S. Trost, G. Oberlender, Predicting accuracy of early cost estimates using factor analysis and multivariate regression, J. Constr. Eng. Manage. ASCE 129 (2) (2003) 198–204.
[18] B. Tysoe, Construction Cost and Price Indices: Description and Use, University Press, Cambridge, 1981.
[19] C. Wilmot, B. Mei, Neural network modeling of highway construction costs, J. Constr. Eng. Manage. 131 (7) (2005) 765–771.