Construction Cost Predication Model Using Macro Economic Indicators

Craig Capano, Ph.D., Jeanette Hariharan, Ph.D., and Hashem Izadi Moud, Ph.D.
Florida Gulf Coast University
Fort Myers, Florida

Ashish Asutosh, Ph.D. Candidate
University of Florida
Gainesville, Florida

Estimating future costs of construction is an important component to the success of any contracting company. Traditionally a cost modifier has been utilized to offset cost escalations or volatility predictions. Construction estimators and contractors have also attempted to utilize a variety of prediction models. This paper establishes a basis for reliable forecasting and explores the possibility of developing prediction models using time series Neural Networks (NN) by utilizing historic data of three accepted macro-economic composite indicators (MEI) and two accepted construction industry cost indices. The use of these macro-economic indicators for NN-based models may be used to predict cost escalations for construction. Nonlinear autoregressive NN models are constructed through using the macro-economic data and the construction cost data to determine if a reliable time-series predictive model could be established. The results of these models indicated that there is a high correlation between the macro-economic escalations, independent factors, and the construction cost escalations, dependent factors, over time. Use and knowledge of these correlations could aid in the prediction of cost escalations during construction.

Key Words: Construction forecast, estimating, economic indicators, neural networks

Introduction

Forecasting construction project cost is a key to the success of any contracting company. Investment decisions can be supported by accurately forecasting construction costs and can allow maximizing profits at all stages of the project (Kong, 2008). The possibility of radical change in cost during the life span of a construction project can cause uncertainty for its financial success (Blair, 1992). The causes of construction cost escalation are several and complex (Yeo, 1990). These include labor, material and equipment cost inflation, taxes, market vulnerabilities, demand and governmental influence. The lifeblood of a construction company is the contractor's expertise in adequate quantity surveying and precise costs estimates. Comprehensive quantity surveys, quotes from vendors, overhead, contingency and profit are mostly volatile and unknown. This occurs due to various reasons such as incomplete plans, price inflation of materials and labor over the life cycle of the project (Capano, 2003). It is also a concern that inflation sometimes is being used as an excuse for poor management practices and inversely, is frustrating for project managers because it has been almost

implausible to identify cost escalation effects to any degree of accuracy (Bassili, 1978). Typically, a contractor uses a cost adjustment technique known as a cost multiplier which is stochastic in nature and is based on experience. There are models which have been developed to help predict future costs by using historical data and applying mathematical concepts (Ostwald, 2000). But the variables which affect these cost escalations can be labyrinthine in nature and are generally based on judgement and industry experience. Many studies have been conducted to create valid models using time series, neural networks, and linear regression but mostly lack accuracy in results (Capano, 2003). Forecast models have been created by narrowing the scope in the construction industry, for example, focusing on highway construction but it was not accurate in estimating different elements (Herbsman, 1986). It was found that the inaccuracy with forecasting issue could be minimized or even solved by introducing an "escalation clause" which is mutually agreed in the contract documents. This clause is not commonly used in North America (Blair, 1992). However, it is important to recognize that no forecast model can provide absolute accuracy with no knowledge of future. Global construction market is predicted to grow over 24.3 trillion U.S. dollars. This is an alarming state where investors and consumers need to be protected on their investments where the growth of the industry is fast pace but there is no accurate way of predict future project costs over time. The construction industry is not the only industry that needs the competency to successfully predict the cost escalation of materials and labor. Other industries often use the economic indicators to develop business strategies, goals and to determine their product prices. While many forecast models are developed using different factors such as inflation, labor cost, equipment cost, none provided a reasonable accuracy in forecasting construction industry future in macro level. This paper (1) explores the possibility of developing forecast models using Neural Networks (NN) by using historic data of three economic composite indices and two construction industry cost indices and (2) propose different NN-based prediction models that connect macro-economic indicators, as independent factors, to construction industry indicators, as dependent factors (Capano, 2003). There are three composite economic indices that are widely accepted to analyze the direction of economics including (1) Leading Economic Indicator (LEI), (2) Coincident Economic Indicator (CEI), and (3) Lagging Economic Indicator (LagEI) indicators. The two construction cost indices that also used in this research are: (1) Construction Cost Index (CCI) and (2) Building Cost Index (BCI).

Construction Economy Cost and Macro-Economic Composite Indices

General construction economy costs indices including CCI and BCI from ENR (2019) and macro-economic indicators LEI, CEI and LagEI from *The Conference Board* (conference board, 2019) were gathered for a period of 41 years on a quarterly basis ranging from September 1978 to June 2019; shown in Tables 1, 2 and 3. This time frame was selected to reflect a myriad of economic upturns, downturns, political influences, recessions, and business cycles. ENR (2019) publishes two indexes that reflect the cost of construction in the United States.

Table 1. LEI Composite

(LEI) Leading Index Composite	Standardization Factor
1. Average weekly hours, manufacturing	27.95%
2. Average weekly initial claims for unemployment insurance	3.24%
3. Manufacturers' new orders, consumer goods and materials	8.32%
4. ISM® new orders index	15.86%
5. Manufacturers' new orders, non-defense capital goods	4.05%
6. Building permits, new private housing units	2.90%
7. Stock prices, 500 common stocks	3.95%
8. *Leading Credit Index*™	8.13%

9.	Interest rate spread, 10-year Treasury bonds less federal funds	11.32%
10.	Avg. consumer expectations for business conditions	14.28%
	Total	100%

CCI is built by combining 200 hours of common labor at the 20-city average of common labor rates, plus 25 cwt of standard structural steel shapes at the mill price prior to 1996 and the fabricated 20-city price from 1996, plus 1.128 tons of Portland cement at the 20-city price, plus 1,088 board-ft. of 2 x 4 lumber at the 20-city price. The second indicator, BCI is determined by combining 66.38 hours of skilled labor at the 20-city average of bricklayers, carpenters and structural ironworkers' rates, plus 25 cwt of standard structural steel shapes at the mill price prior to 1996 and the fabricated 20-city price from 1996, plus 1.128 tons of Portland cement at the 20-city price, plus 1,088 board-ft. of 2 x 4 lumber at the 20-city price. LEI, CEI and LagEI macro-economic indicators continue to be the key elements in an analytic system designed to signal peaks and troughs in the business cycle. Because they are averages, they tend to smooth out a good part of the volatility of the individual series and thereby serve as summary measures of the business cycle. The LEI index components reflect the degree of tightness in the labor market due to employer hiring and firing; the buildup of orders and contracts that effect future production; materials prices that reflect shortages or gluts of raw materials used to expand or reduce existing inventories; financial conditions associated with the availability of funds in credit markets; and consumer psychology that effects household spending. This index can also be thought of as a gauge for the future state of the economy. Since it is comprised of several factors such as new orders, new housing, and consumer expectations, it can be used to analyze the near trend in economic output. Table 1 contains the index of leading economic indicators, which is a composite of the following ten specific indicators and their standardization factors. Table 2 CEI index components reflect employment, real incomes generated from production, output in cyclically sensitive mining and manufacturing industries, and real manufacturing and trade sales depicting the flow of goods. This index can be thought of as a reflection of the current state of the economy.

Table 2. CEI Composite

	(CEI) Coincident Index	Standardization Factor (weighting)
1.	Employees on nonagricultural payrolls	52.90%
2.	Personal income less transfer payments	20.52%
3.	Industrial production	14.54%
4.	Manufacturing and trade sales	12.02%
	Total	100%

The last composite index utilized by economists to determine the trend in business is the LagEI index. This index's components reflect: the effect of the duration of unemployment on business costs of recruitment and training, the cost of maintaining inventories, labor cost per unit of output, the burden of paying back business loans, interest payments as cost of production and prices of consumer services. Lagging index can be used as a tool to analyze the future trend in economy. Since it is comprised of indicators that could have an effect of the economy at a future time in Table 3.

Table 3. LagEI Composite

	(LagEI) Lagging Index	Standardization Factor
1.	Inventories to sales ratio, manufacturing and trade	12.70%
2.	Average duration of unemployment	3.69%
3.	Consumer installment credit to personal income ratio	18.24%
4.	Commercial and industrial loans	9.33%

5.	Average prime rate	30.15%
6.	Labor cost per unit of output, manufacturing	5.05%
7.	Consumer price index for services	20.84%
	Total	100%

Historically, the cyclical turning points in the leading index have occurred before those in aggregate economic activity, while the cyclical turning points in the coincident index have occurred at about the same time as those in aggregate economic activity. The cyclical turning points in the lagging index generally have occurred after those in aggregate economic activity. A change in direction in a composite index does not signal a cyclical turning point unless the movement is of significant size, duration, and scope. It is important to recognize that the timing of the leading index has been irregular and "false signals" are inevitable. The main value of the leading index is in signaling that either the risk of a recession has increased or that a recession may be coming to an end. The leading, coincident, and lagging indexes are an analytical system of assessing current and future economic trends, particularly cyclical expansions and recessions (Frumkin, 1994). Figure 1 illustrates all the data used in this research including CEI, LEI, LagEI, CCI and BCI, from 1978 to 2019.
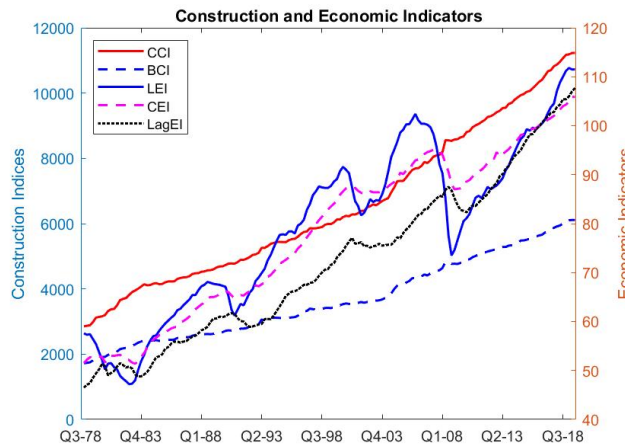

Figure 1. Building and Construction Cost indices corresponding to Economic Indicators

Methodology

Data beginning in December of 1978 were recorded and updated through the period of the second quarter of 2019. There is one outstanding epoch period to be noted that occurs in the data for the LEI and CEI starting from March 2008 and continuing through March 2011. The CCI remained steady during this period, and its trend is seen to have predicted the economic recovery. Citing this data, it will be interesting to discern if: (1) there is a distinct correlation between the CCI and BCI to the macro-economic factors of LEI, CEI, and LagEI using time-series analysis; (2) Trends can be predicted for future economic assessments based on the time-series information between CCI-BCI and the macro-economic factors data; (3) the error associated with these predictive analysis models is reasonable, so that the model can be used reliably for future economic forecasts.

Static Linear Regression to Establish Correlation

Several analysis tools are used in this study. Static linear regression analysis is used first, to understand if a general pattern could be found to fit the CCI and BCI as dependent variables to the economic forecast model. A stationarity test is used to establish the regression model validity. A curve fit is applied to the data for a BCI and CCI verses LEI-LagEI-CEI data sets to prove stationarity of data as required for static regression for time series data (Nielsen, 2005). The polynomial fit was then differentiated using first order central differencing approximation and the results are shown in Figure 2. The initial results from this first method of analysis showed that the BCI data was correlated to an R-squared value of 96.66% for the LagEI, 88.06% for CEI, and 74.21% for the LEI. Since the R-squared values are reasonable to assume this correlation existed, and there is indeed a strong economic correlation with the BCI and CCI parameters, another tool is used to gain insight on the general dependence formula that could be derived and to further validate our findings from our static estimation relationship.
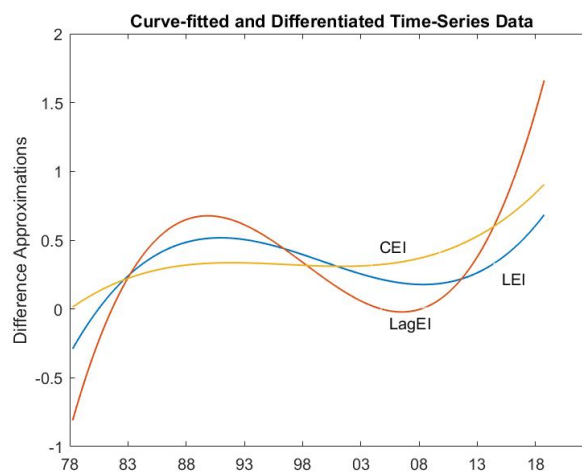

Figure 2. Stationarity Proof for Regressions Analysis of LEI, CEI and LagEI

Nonlinear Autoregressive Neural Network

A modified Levenberg-Marquardt NN was used to estimate a polynomial fit and to optimize the function representation for each of the construction indices to their dependent variables, LEI, CCI and LagEI. This model is based on the Group Method of Data Handling (GMDH). The GMDH is a forecasting type NN whose algorithm is a self-organizing method for complex systems modelling. These GMDH forecasters are used for understanding and predicting outcomes for physical laws, approximation of data in medical studies, pattern recognition, economic estimates, and extrapolation of field theory data. The basic premise of the GMDH model is developed from linear algebra and reconstructs polynomial fits of the data. The heuristic and self-organizing nature of the algorithm rely heavily on constructing higher order polynomials for localization, and using feedback criteria, and error analysis for prediction. GMDH is used extensively for modelling time series data. The basic statistical models are not always enough for interdependent variable data. Since our model for predictive analysis is ordered by several indices, time series and economic factors, the GMDH NN estimate was used to give a better prediction of how the construction market indices are influenced by the economic factors. The GMDH algorithm uses all pairwise combinations of *p* lagged time series. Each combination enters each neuron. Using our two inputs a model is constructed that predicts the desired output. These two input variables are inputs to a neuron, and one result goes to an output. The structure of the model is specified by (Song, 1027). The specification requires that six coefficients (or

weights) in each model are to be estimated, for our case with two input variables and each delayed. The polynomial equation derived for two inputs for each layer is given by (1):

$$y = a + \sum_{i=1}^{m} b_i x_i + \sum_{i=1}^{m} \sum_{i=1}^{m} c_{ij} x_i x_j \qquad \textbf{(1)}$$

Where m is the number of variables and a, b_i, c_{ij} are the coefficients of the variables in the polynomial. The output, y, is the response variable; x_i and x_j are the lagged time series to be regressed.

Data Analysis and Results

Results from GMDH Model for time-series prediction

The time-series projection model based on the GMDH architecture described was used to develop a time-based polynomial mapping of the data, and to provide a prediction model for forecasting future economic trends in correspondence with the BCI-CCI matrix. The time-line is noted by "epochs" which are time periods of four years starting from the fourth quarter of 1978. Using two different tools for this GMDH model, a time-series prediction model both in MATLAB and GMDH gave similar results. The LEI model shows errors in the economic indicators around the year 2008-2009. The models show that they accurately can fit time series data of the BCI and CCI data and map these as parameters into the economic indicators. The models were validated (green data points on graphs) to give predictive indicator results that closely followed the actual data. The red data points were test results from adding randomized data points to the original data and then testing the model derived auto-regression function. Using the models to predict for more than 12 months beyond the June 2019 quarter (Figure 5), showed the accuracy of using these models for prediction (i.e. error is less than ~2%). All models indicate a slowing about 20 months past the June 2019 quarter, and one of the models suggested a minimum economic indicator period occurring between the third quarter of 2021 and 2023. These graphs are shown in comparison analysis in Figures 3 through 5. The error, correlation and R-squared values are shown in the Table 4 for the GMDH, BCI model. All the results showed similar, accurate and reliable model statistics for the time-series modeling of the data, and time-series prediction.
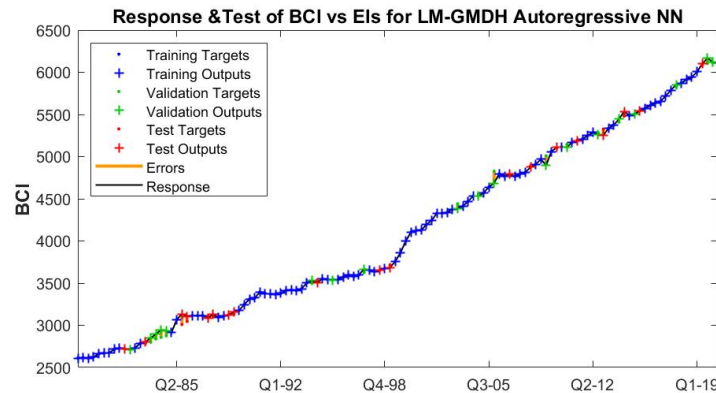

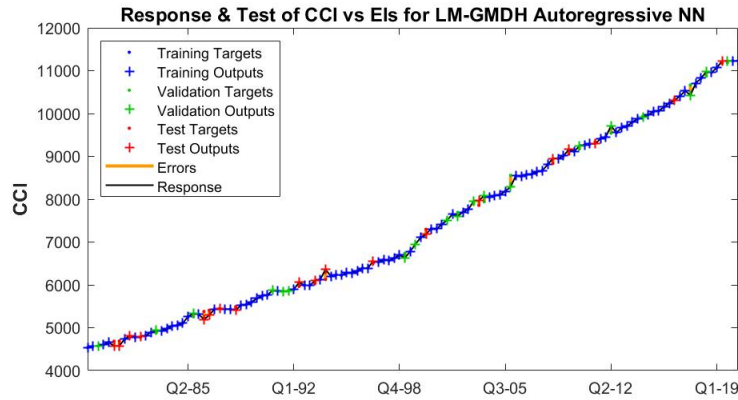Figure 3. GMDH Time-Series Model for BCI vs. LEI, CEI, LagEI

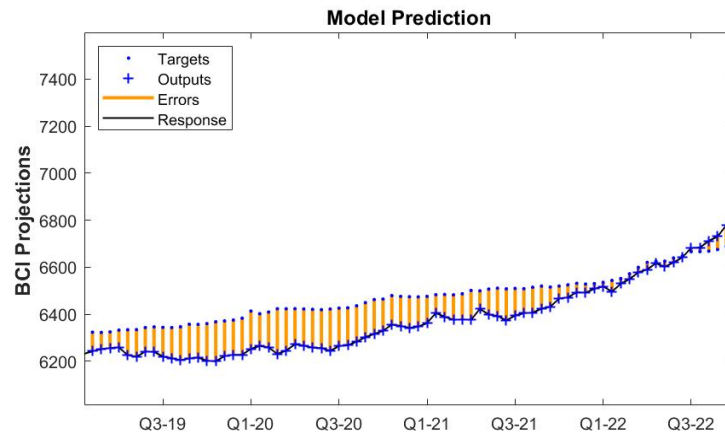Figure 4. GMDH Time-Series Model for CCI vs. LEI, CEI, LagEI

Figure 5. BCI Projections based on LEI, CEI, LagEI from
Nonlinear Autoregressive GMDH Model

Table 4 Error, Correlation and R-squared Values for Test Data

Statistics for the BCI Model (GMDH)	Model Fit
Mean Square Error (MSE)	3447.525
Standard Deviation of Residuals	131.56
Coefficient of Determination (R^2)	0.907
Correlation	0.827

Discussion and Conclusions

This study is designed to discover if: (1) There is a distinct correlation between the CCI and BCI to the macro-economic factors, (2) Trends can be predicted for future economic assessments based on the time-series auto-regression NN predictive model, which verifies correlation between CCI-BCI and the macro-economic factors data and (3) The error associated with these predictive analysis models is reasonable, so that the model can be used reliably for future economic forecasts.

The approach, methodology and results, have shown with clear, consolidated evidence, that all three of these projections are indeed indicated. The linear regression and NN auto-regressive fit models pointed to a strong relationship between CCI-BCI and economic indicators. We also discovered that there is a strong inverse relationship of MEI to CCI-BCI. Using the Levenberg-Marquardt NN with GMDH technique for optimal fit and time-series modeling and prediction, all economic indicators were shown to be correlated strongly enough to build a reliable predictive model. All error was well within the bounds of normal tolerance and the R-squared values were reasonable, and in most cases above what is considered acceptable for dependable prediction models. Further validation and testing of these models are underway and as more economic and construction data is added to these models, more accurate stochastic methods can be determined and these models updated.

References

Bassili, A. R. & Finlayson, H. W. (1978). Escalation. *Project Management Quarterly, 9*(3), 27–30.

Blair, A. N., Lye, L. M., & Campbell, W. J. (1993). Forecasting construction cost escalation. *Canadian Journal of Civil Engineering*, *20*(4), 602-612.

Capano, C, and Saeed Karshenas. (2003) "Applying accepted economic indicators to predict cost escalation for construction." *ASC Proc., 39th Annual Conf.* 2003.

Conference-board (2019). Available: <https://www.conference-board.org/us>

ENR, (2019). Engineering News Record Journal. Available: < https://bit.ly/35IxZnd >

Frumkin, N. (1994). Guide To Economic Indicators, second edition, M.E. Sharpe

Herbsman, J. (1986). Model for forecasting highway construction cost. *Transportation Research Record*, *1056*, 47-54.

Hendrickson, C., Hendrickson, C. T., & Au, T. (1989). *Project management for construction: Fundamental concepts for owners, engineers, architects, and builders*. Chris Hendrickson.

Kong, F., Wu, X. J., & Cai, L. Y. (2008). A novel approach based on support vector machine to forecasting the construction project cost. In *2008 International Symposium on Computational Intelligence and Design*. IEEE.

Nielsen, H.B. (2005, September). Linear Regression with Time Series Data. In Econometrics 2, Lecture Notes, University of Copenhagen.

Sinha, Sunil K.; McKim, Robert A. (1997). "Forecasting Construction Cost Escalation Using Artificial Neural Networks," *Artificial Neural Networks in Engineering Conference*.

Song, S., Wang, L. (2017, November). "Modified GMDH-NN algorithm and its Application for Global Sensitivity Analysis", *Journal of Computational Physics,* Vol. 348, pp. 534-548.

Yeo, K.T. (1990). "Risks, Classification of Estimates, and Contingency Management," *Journal of Management in Engineering*, ASCE 10(1), p 72-84.