Forecasting Construction Cost Index using Energy Price as an Explanatory Variable

Variations of ENR (Engineering News Record) Construction Cost Index (CCI) are problematic for cost estimation and bid preparation. Accurate forecasting of CCI is important for cost estimation and budgeting. The Objective of this study is to create a proper multivariate time series model including CCI and an energy price index for forecasting CCI accurately. Statistical tests verify the relevance and examine the characteristics of the energy price index. Based on the results of the statistical tests, an appropriate multivariate time series model is created to forecast CCI. It is shown that CCIs predicted by the multivariate model are more accurate than those predicted by the seasonal autoregressive integrated moving average model and Holt-Winters exponential smoothing model that are two most accurate univariate time series models. It is expected that the proposed model assist cost engineers and investment planners to prepare more accurate cost estimates and budgets for capital projects.

Key Words: ENR Construction Cost Index, Energy Price, Forecasting

Introduction

Construction Cost Index (CCI) has been published by Engineering News-Record (ENR) monthly in the United States. It has broadly been utilized for estimating costs and preparing bids in the United States. ENR defines Construction Cost Index (CCI) as the weighted average price of common labour, standard structural steel, Portland cement and lumber in 20 cities (ENR, 2011). Historical costs are typically used for construction cost estimation. It is crucial to consider the construction cost changes over time (Hendrickson, 1998). Cost estimators usually take into account the changes in construction cost by forecasting future cost indices and correcting the original cost estimates accordingly. Accurate prediction of CCI can result in preparing accurate bids and prevent under- or over-estimation. Variations of CCI are challenging for cost estimation and bid preparation.

The objective of this research is to create appropriate multivariate time series model including CCI and an energy price index for predicting the variations of CCI accurately. We propose multivariate time series model to forecast CCI because it uses the information in two time series (CCI and an energy price index) to forecast CCI and it can potentially be more accurate than univariate time series models. Our hypothesis is that energy price level contains information that is useful to predict CCI.

In the rest of this section, the research background on CCI forecasting is reviewed. In Research Method section, statistical tests (Pearson correlation test, unit root test, Granger causality test and Johansen's cointegration test) are utilized to verify the relevance and to examine the characteristics of the energy price index and CCI. A multivariate time series model is proposed at the end of Research Method section. The Results section presents the results of statistical tests, the estimation of the proposed multivariate time series model and prediction of CCI using the proposed model. Conclusions are presented in the last section.

Two categories of quantitative methods have been proposed to forecast construction cost indices (Touran and Lopez, 2006): Causal and statistical methods. The statistical methods use time series analysis and curve fitting to forecast construction cost indices (Hanna and Blair, 1993). (***) compared various univariate time series models to forecast CCI and concluded that seasonal autoregressive integrated moving average model and Holt-Winters exponential smoothing are the most accurate univariate time series approaches for forecasting of CCI. The major problem of statistical methods is that they do not have explanatory capability and they are just suitable for short-term forecasting (Touran and Lopez, 2006; Goh and Teo, 2000). The causal methods forecast construction cost index based on its explanatory relationship with other variables. Williams (1994) used the trends in CCI, prime lending rate, housing starts, and the months of the year as the inputs of back-propagation network models to predict CCI. He concluded that CCI prediction is a complex problem and CCI cannot be accurately predicted by the neural network models.

Energy price level is widely ignored as one of the potential explanatory variables of construction cost. Our hypothesis is that energy price level can be used along with CCI within a proper multivariate time series model to forecast CCI more accurately than the existing univariate time series models. Crude oil price is used in this study as a measure for representing energy price level.

Research method

Statistical tests

Four statistical tests are used to verify the relevance and examine the characteristics of crude oil price as the explanatory variable of CCI. Pearson correlation analysis is used to study the linear relationship between crude oil price and CCI. The test statistic of this test is based on Pearson product-moment correlation coefficient and the null hypothesis of the test is that there is no association between CCI and crude oil price.

Time series tests are preceded by another test for identifying the integrated order of the variables. Augmented Dickey-Fuller (ADF) test proposed by Dickey and Fuller (1979) and extended by Said and Dickey (1984) is used in this study for identifying the order of integration of crude oil price and CCI. The null hypothesis is that the time series under study is not stationary and the alternative hypothesis is that the time series is stationary. A time series is stationary if its statistical properties do not change after being time-shifted (Brockwell and Davis, 2002). Critical values recommended by Banerjee et al. (1993) are used for the unit root test. Akaike Information Criterion (AIC) is used to identify the lag lengths (Akaike, 1974). If the results of the unit root tests show that the variable under consideration is not stationary, the time series data are differenced and the unit root test are repeated for the differenced terms. The number of times that a variable is differenced until stationary time series data is created is called the "order of integration" for the variable.

Grangter causality test is the third statistical test used in this study. Granger causality test is a statistical test to determine whether the information in time series of one variable is useful to predict the value of another variable (Granger, 1969). "*X* Granger causes *Y"* means that past values of *X* are useful to predict the future values of *Y*. The null hypothesis of the test is that the past *p* values of *X* do not help the predictability of *Y*. Since the Granger causality test is sensitive to the number of lags (*p*), it is applied for 6, 12, 18, 24, 30 and 36 lag lengths to examine if the information in time series of crude oil price is useful for predicting future values of CCI. These lag lengths represent a 3-year time horizon.

The fourth statistical test used in this study is the cointegration test. Cointegration test is used to examine if the CCI and crude oil price are cointegrated. This test is important since it has key impact on the selection of appropriate multivariate time series models. In this research, a cointegration test proposed by Johansen (1988) and extended by Johansen and Juselius (1990) is used to examine whether CCI is cointegrated with crude oil price. The null hypothesis of this test is that the number of cointegrating relationship is equal to zero (No cointegration). Akaike Information Criterion (AIC) is used to specify the lag length for the cointegration analysis in this study (Akaike, 1974).

Multivariate time series model including CCI and Crude Oil Price

Vector Error Correction (VEC) models are recommended for time series modelling where variables are cointegrated (Pfaff, 2008). The following form of VEC models is used in this study:

$$\Delta y_t = \sum_{i=1}^{p-1} A_i \Delta y_{t-i} + B y_{t-p} + C + \varepsilon_t$$

where y_t is the (2×1) vector of time series at period t, A_i (i=1,…,p-1) are (2×2) coefficient matrices of endogenous variables, B is (2×2) coefficient matrix containing the cumulative long-run impacts, C is (2×1) vector of constants, and εt is (2×1) vector of error terms. Gaussian maximum likelihood (ML) procedure (Johansen, 1995) is used to estimate the coefficients of the VEC model.

Empirical Results and Discussions

Data

We used the ENR monthly CCI data from January 1975 to December 2008 as CCI time series data for conducting the statistical tests. Crude oil price is also available for this period. Data about crude oil price is collected from U.S. Energy Information Administration. Crude Oil Price is the domestic first purchase price of a barrel of crude oil. Training data from January 1975 to December 2008 is used for estimating the parameters of the multivariate time series model. Testing data from January 2009 to December 2011 is used to assess the predictability of the multivariate model.

Results of Statistical tests

Based on the correlation result, crude oil price (+0.73) is significantly correlated with CCI at 1% significance level. This result shows that there is a strong positive linear relationship between CCI and crude oil price. The integrated orders of crude oil price and CCI are determined using ADF unit root tests. The results of ADF unit root tests are presented in Table 1. Table 1 shows that CCI and crude oil price are not stationary. CCI and crude oil price become stationary by applying the differencing operator once. The null hypothesis of non-stationarity is rejected at 1% significance level for both differenced CCI and differenced crude oil price. Therefore, CCI and crude oil price are integrated of order 1.

Table 1
Results of ADF unit root tests for CCI and crude oil price

Variable	**ADF t-statistic**	**Variable**	**ADF t-statistic**
CCI	0.70[6]	ΔCCI	-8.70*[5]
COP	-2.08[6]	ΔCOP	-5.36*[10]

Notes: Δ is the first difference operator; * Rejection of the null hypothesis at the 1% significance level; [.] denotes the lag order that is selected based on the AIC criterion;

Bivariate regression models are used to test the Granger causality relationship between crude oil price and CCI. Table 2 summarizes the results of Granger causality tests between CCI and crude oil price. The results summarized in Table 2 show that crude oil price consistently Granger cause CCI at all specified lag lengths. Therefore, historical values of crude oil price are useful to predict CCI. Although Granger causality test shows that historical values of crude oil price are useful to predict CCI, it does not specify the type of multivariate time series model that can be used to take advantage of the historical information. Cointegration test is used to identify the proper type of time series model for characterizing the variations of CCI using information in historical time series of CCI and crude oil price.

Table 2
Results of Granger causality test between CCI and crude oil price

Null hypothesis	**F Statistics**					
	Lag 6	**Lag 12**	**Lag 18**	**Lag 24**	**Lag 30**	**Lag 36**
ΔCOP does not Granger cause ΔCCI	6.42*	3.78*	2.86*	2.40*	2.21*	2.19*

Note: Δ is the first difference operator; * Rejection of the null hypothesis at the 1% significance level.

The results of Johansen's cointegration test for the bivariate vector of CCI and crude oil price are presented in Table 3. The trace statistics show that null hypothesis of no cointegrating relationships between CCI and crude oil price can be rejected at 1% significant level. Therefore, CCI and crude oil price are cointegrated at the 1% significance level. This cointegration has implication on the selection of appropriate type of time series model for characterizing the variation of CCI.

Table 3
Results of the Johansen cointegration tests

Explanatory Variables	Lag length	Trace statistics (H0: r=0)	Trace statistics (H0: r≤1)
CCI & COP	6	64.86*	6.65

Multivariate time series model

Vector Error Correction (VEC) models are recommended for multivariate time series modelling where the variables are cointegrated (Pfaff 2008). Since CCI and crude oil price are cointegrated, we created a VEC model using CCI and crude oil price to characterize variation of CCI. The coefficients of this VEC model are estimated using Gaussian maximum likelihood procedure (Johansen 1995).

The residuals of the VEC model should not have serial correlation and their variance should be constant (Lack of heteroskedasticity). Existence of serial correlation among the residuals of the VEC model is diagnosed using Breusch-Godfrey LM test proposed by Breusch (1978) and Godfrey (1978). ARCH test (Engle, 1982) is applied for investigating heteroskedasticity in the residuals of CCI in the VEC model. The Breusch-Godfrey and ARCH test statistics are 30.3 and 11.8 for our VEC model, respectively. These test statistics do not lead to the rejection of the null hypotheses of Breusch-Godfrey and ARCH tests. Therefore, the VEC model passes both diagnostic tests.

The predictability of the VEC models is compared with the predictability of two existing univariate time series models for forecasting CCI. (***) studied univariate time series models and proposed seasonal autoregressive integrated moving average model and Holt-Winters exponential smoothing model as the most accurate univariate time series approaches for forecasting CCI. These two models are recreated in this study to be compared with the multivariate time series model in terms of the predictability. The predictability of the time series models are compared based on Mean Absolute Prediction Error (MAPE) and Mean Squared Error (MSE). Table 4 presents MAPE and MSE calculated using forecasted data points from the models and the testing data from January 2009 to December 2011. Based on the results shown in Table 4, the VEC model provides better forecasts than the univariate time series models.

Table 4
Forecasting errors of time series models

Measure	VEC	Seasonal ARIMA	Holt-Winters Exponential Smoothing
MAPE	0.96%	1.40%	2.68%
MSE	10544.9	17921.6	86890.7

Conclusions

Accurate prediction of CCI is important for cost estimation and budgeting of capital projects. A multivariate time series model is created to forecast CCI accurately using the historical information in CCI and crude oil price. Statistical tests (Pearson correlation test, unit root test, Granger causality test and cointegration test) are used verify the relevance and examine the characteristics of crude oil price as an explanatory variable for CCI. Statistical tests show that crude oil price is a consistent explanatory variable of CCI. They also show that it is cointegrated with CCI. Based on the results of the results of the cointegration test, a Vector Error Correction (VEC) model is created to forecast CCI. The coefficients of the VEC model are estimated using Gaussian maximum likelihood procedure. The VEC model is diagnosed using two statistical tests (Breusch-Godfrey test and ARCH test). The results of the diagnostic tests show that the VEC model passes both diagnostic tests. The predictability of the VEC model is compared with seasonal autoregressive integrated moving average model and Holt-Winters exponential smoothing model. The VEC model provides more accurate forecasts than the univariate time series models. It is anticipated that the proposed VEC model help cost engineers and capital planners prepare more accurate bids, cost estimates, and budgets for capital projects.

References

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19 (6): 716–723.

Banerjee, A., Dolado, J.J., Galbraith, J.W. & Hendry, D.F. (1993). Cointegration, error correction, and the econometric analysis of non-stationary data. Oxford University Press, Oxford.

Breusch, T.S. (1978). Testing for autocorrelation in dynamic linear models. Australian Economic Papers, 17, 334-355.

Brockwell P.J. and Davis, R.A. (2002). Introduction to time series and forecasting. Second Edition. Springer-Verlag, New York.

Dickey, D.A. & Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with unit roots. Journal of the American Statistical Association, 74, 427-431.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50, 987-1007.

ENR (2011). Engineering News Record. http://enr.construction.com/economics/, Last visit: Sep/20/2012.

Godfrey, L.G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. Econometrica, 46: 1303-1313.

Goh, B.H. & Teo, H.P. (2000). Forecasting construction industry demand, price and productivity in Singapore: the Box-Jenkins approach. Construction Management and Economics, 18, 607-618.

Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 37(3), 424-438.

Hanna, A.S., and Blair, A. N. (1993). Computerized approach for forecasting the rate of cost escalation. Proceeding of Computing in Civil and Building Engineering Conference, 401-408.

Hendrickson, C. (1998). Project management for construction. Available online at http://pmbook.ce.cmu.edu/ , last visit: Dec/08/2011.

Johansen, S. (1995). Likelihood-based inference in cointegrated autoregressive models. Oxford University Press, Oxford.

Johansen, S. (1988). Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control, 12, 231-254.

Johansen, S., and Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration-with applications to the demand for money. Oxford Bulletin of Economics and Statistics, 52(2), 169-210.

Pfaff, B. (2008). Analysis of integrated and cointegrated time series with R. Second Edition, Springer, New York, NY.

Said, E.S. & Dickey, D.A. (1984). Testing for unit roots in autoregressive average models of unknown order. Biometrika, 7(3), 599-607.

Touran, A. and Lopez, R. (2006). Modelling Cost Escalation in Large Infrastructure. Construction Engineering and Management, ASCE, 132(8), 853-860.

Williams, T.P. (1994). Predicting changes in construction cost indexes using neural networks. Journal of Construction Engineering and Management, ASCE, 120(2), 306-320.