Exponential Smoothing Time Series Models for Forecasting ENR Construction Cost Index

B. Ashuri, PhD and J. Lu
Economics of the Sustainable Built Environment (ESBE) Lab
School of Building Construction, Georgia Institute of Technology
280 Ferst Drive 1st Floor Atlanta, GA 30332-0680

Every month, Engineering News-Record (ENR) publishes Construction Cost Index (CCI), which is a weighted-aggregate index of the 20-city average prices of construction activities. Although CCI is increasing in the long-term, it is subject to considerable short-term variations, which make it problematic for cost estimators to prepare accurate estimates. The ability to predict CCI can result in more-accurate bids and avoid under- or over-estimation. Time series analysis is used to develop CCI forecasting models based on two Exponential Smoothing (ES) methods: Holt ES and Holt-Winters ES. ES time series methods investigate the underlying characteristics of the CCI data and makes systematic forecasts. The predictability of both developed ES models is better than the predictability of the ENR subject matter experts' forecasts. The developed forecasting models can be used to prepare more-accurate estimates for contractors and budgets for owners and reduce construction costs by better-timed project execution.

Keywords: ENR, Construction Costs Index, Exponential Smoothing (ES), Forecast

Introduction

Engineering News-Record (ENR) monthly publishes the Construction Cost Index (CCI), which is widely used by cost estimators in America to estimate construction project costs or prepare budgets for owner organizations. However, historical ENR CCI data shows considerable short-term variations while increasing in the long-term, as shown in Figure 1. Variation of CCI will severely affect the accuracy of the estimation of construction cost or budget. Recent study shows variability in construction cost is one of the most important risks impacting the contractor's profit and owner's budget (Ervin 2007; Gallagher 2008). Additionally, contractors will suffer higher construction costs, delayed progress or even insolvencies accompanied with high variation of CCI. In consequence, owners will also take corresponding risks transferred by constrictors, such as higher bids, large hidden contingencies or short bid guarantees.

In order to avoid the loss caused by variation of CCI, a time series method will be applied to develop a forecast model to provide more accurate CCI prediction. The research process would be following: section 2 provides literature review in prediction of cost index; section 3 investigates time series properties of CCI data with respect to stationary and seasonality; section 4 applies ES time series approaches to fit historical CCI dataset and conducts diagnostic analysis to test the model applicability; section 5 compares time series model and ENR's subject matter experts' CCI forecast with respect to predictability; section 6 states the conclusion of paper.

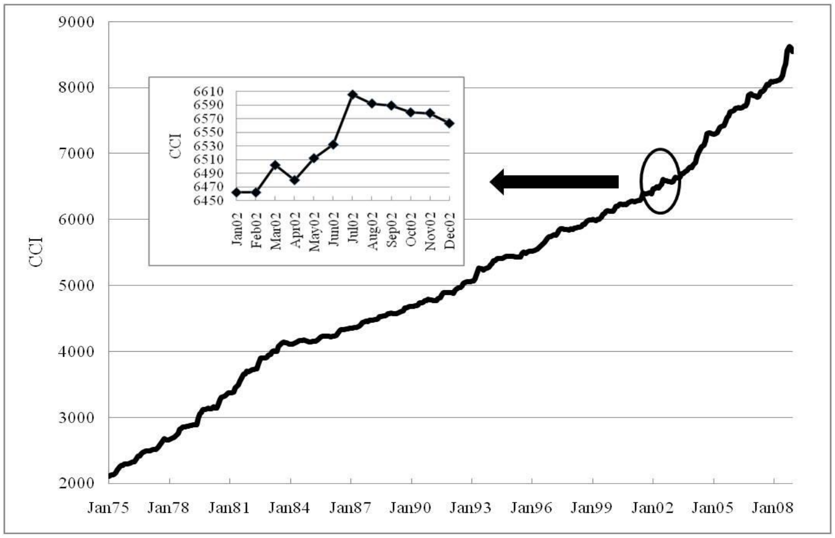

Figure 1: The long term increasing trend and short-term variations of CCI data

Research Background

There are two categories of predictive modeling in previous works (Taylor and Bowen, 1987): the causal method and the time series method. The causal method assumes that the predicted variable is determined by independent explanatory variables, i.e. regression. The causal method has been applied to predict tender price indexes (Akintoye and Skitmore 1994; McCaffer et al. 1984), building price (Runeson 1988), construction cost (Koehn and Navvabi 1989), early cost estimation (Trost and Oberlender 2003), and construction labor cost in Hong Kong (Wong, 2005).

On the other hand, the time series method determines future trends based on past values and corresponding errors. Since a time series method only require the historical data of forecast variable itself, it is widely used to develop predictive models. The time series method has been used to forecast Taiwan's construction cost indexes (Wang and Mei 1998), property prices (Chin and Mital, 1998), building costs (Taylor and Bowen 1987), and tender price index (Fellows 1991; Ng et al. 2000). In a rare study of ENR CCI by (Williams 1994), time series method was compared with linear regression and neural network models with respect to predictability.

Following Williams study, our research will explore advanced ES models to forecast CCI and investigate forecasting accuracy of ES model with comparison to ENR's subject matter expert forecasts.

Time Series Properties of ENR CCI Data

The historical ENR CCI dataset $\{I(t); t = 1, 2, \ldots, 360\}$ collected consists of 30 years of monthly data from January 1979 to December 2008. As shown in Figure 1, CCI time series subject to considerable short-term variations despite increasing long-term trends. Seasonality is an important property exhibiting certain cyclical or periodic behaviors in time series data. Generally, a period of 1 year (12 month) is considered for monthly data such as CCI. The box plot in Figure 2 (left) shows that the monthly increment of CCI peak in July, decrease through December, and then begin rising until the next July peak. The AutoCorrelation Function Plot of the monthly increment of CCI in Figure 2 (right) shows that CCI increment has relatively higher autocorrelations at lag levels of multiple 12, which are lags 12, 24, 36, and 48. This is an evidence for the certain seasonality of CCI increment dataset with the period of 12 months (Tsay, 2005).

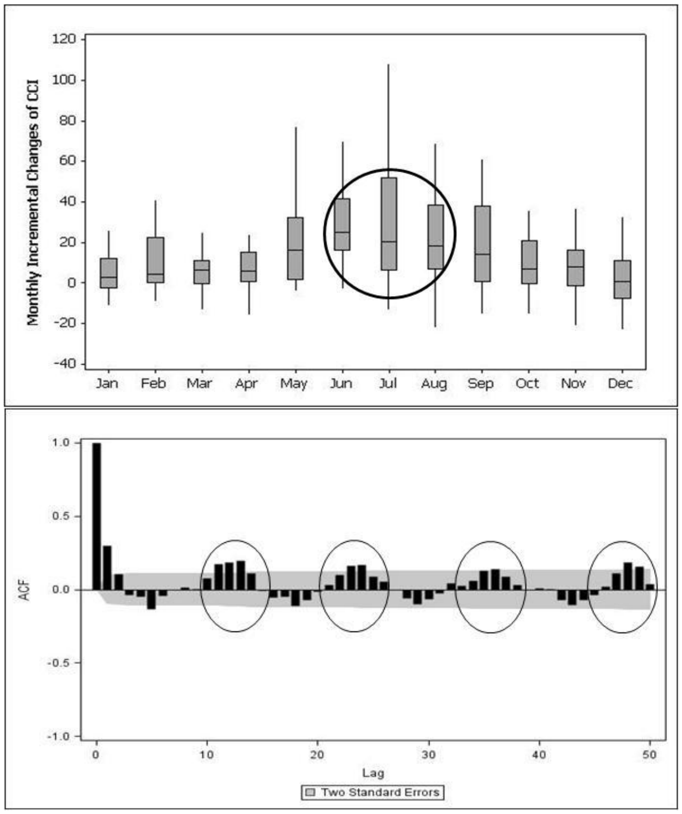

Figure 2. Box plot of monthly increment of CCI

Fitting Exponential Smoothing (ES) Time Series Model

In order to improve predictability, ES models will be fitted to each dataset, then the fitted models are used to make the out-of-sample forecast. The whole dataset from 1979 to 2008 is used as an example to illustrate the model fitting process.

Holt Exponential Smoothing

Holt exponential smoothing (Holt ES) method was developed for the U.S. Office of Naval Research to handle time series datasets with trends (Ord**,** 2004). It is appropriate for time series data that display trends (Brockwell & Davis, 2002). Considering the increasing trend of CCI, shown in Figure 1, the Holt ES model is used for in-sample forecasting. Monthly CCI forecasts are computed based on level and trend smoothing in this approach. Level smoothing estimates the monthly level factor a while trend smoothing estimates the trend factor or the average monthly growth rate (Gardner, 1985). The nature of model building procedure in this approach is recursive as summarized in the following equations.

Level Smoothing: $L(t) = aY(t) + (1-a)\hat{Y}(t)$

Trend Smoothing: $T(t) = b[L(t) - L(t-1)] + (1-b)T(t-1)$

Forecasting: $\hat{Y}(t+1) = L(t) + T(t)$

Where t = 2, 3,…, 360 is the time index; Y(t) is the actual CCI at time t; L(t) is the level factor at time t; T(t) is the trend factor at time t; Ŷ(t) is the forecasted CCI at time t; a and b are constant weighting factors that must be estimated such that the error measure MSE is minimized for this model; The initial values for trend smoothing, level smoothing, and forecasting are T(2)=Y(2)-Y(1), L(2)=Y(2), Ŷ(3)=Y(2)+T(2).

Least Square Estimate (LSE) is used to decide the appropriate values for the weighting factors a and b such that the MSE of the forecasted values is minimized. The optimal values of these parameters for dataset from 1979 to 2008 are summarized in Table 1.

Table 1: Parameter estimation for the Holt ES model

Model Parameter	Estimate (e)	Std. Error (f)	T-Statistic (e/f)	P-value
Level Smoothing Weight (a)	0.99900	0.0382	26.1430	<.0001
Trend Smoothing Weight (b)	0.03741	0.0109	3.4226	0.0007

Table 1 also summarizes the results of statistical t-tests for the significance of parameters a and b in the Holt ES model. The low p-values indicate that both level and trend smoothing weighs are required to build an appropriate fitting model. Error measures of the Holt ES model are shown in Table 2.

Table 2: Measures of error for Holt ES model

Error Measure Type	Error Measure
MAPE	34.75%
MSE	643.56
MAE	18.27

Considering the increasing trends of CCI shown in Figure 1 the Holt ES model is applicable to fit the dataset. The significance of trend smoothing weight or parameter b in the Holt ES model also shows considerable trends in CCI.

Holt-Winters Exponential Smoothing

Winter generalized the Holt ES method to handle the seasonality in time series data by introducing a third factor seasonal smoothing into time series analysis. Seasonal smoothing is an estimated value of seasonal growth rate reflecting the seasonal pattern of time series data (Winter, 1960). Since CCI time series data show certain seasonality, Holt-Winters Exponential Smoothing (Holt-Winters ES) model is fitted to investigate whether the error measures would be improved in comparison with the Holt ES model.

The nature of model building procedure in the Holt-Winters ES approach is also recursive as summarized in Appendix A. Least Square Estimate (LSE) is used to decide the appropriate values for the weighting factors a, b and c such that the MSE of the forecasted values is minimized. The optimal values of these parameters for are summarized in Table 3. Error measures of the Winter-Holt ES model are shown in Table 4.

Table 3: Parameter estimate for Holt-Winters ES model

Model Parameter	Estimate (e)	Std. Error (f)	T-Statistic (e/f)	P-value
Level Smoothing Weight (a)	0.95179	0.0364	26.1419	<.0001
Trend Smoothing Weight (b)	0.04096	0.0113	3.6246	0.0003
Seasonal Smoothing Weight (c)	0.99900	0.7790	1.2824	0.2005

Table 4: The predictability measures for Holt-Winters ES model

Error Measure Type	Error Measure
MAPE	32.96%
MSE	592.96
MAE	17.59

Table 3 also summarizes the results of statistical t-tests for the significance of parameters a, b, and c in the Winter-Holt ES model. The low p-values across three tests indicate that level, trend, and seasonal smoothing weighs are required to build an appropriate fitting model. However, the p-value for seasonality parameter c is larger than 0.05 significance level, indicating that the null hypotheses of c=0 cannot be rejected. Thus, Holt-Winters ES will be reduced to Holt ES on the dataset 1979~2008, although it achieves better error measures with comparison to Holt ES. In the out-of-sample section, the Holt-Winters ES is fitted with different dataset, once it pass the t-test, the Holt-

Winters model for out-of-sample forecasting will be kept; otherwise, reduced to Holt ES model to make the out-of-sample forecasting.

Out-of-sample Forecasting

Out-of-sample forecasting is conducted to make 12-month ahead prediction of CCI, i.e. prediction of CCI for next December based on historical data available till December this year. In order to compare the predictability of ES time series model and ENR Subject Experts' CCI Forecast published in Four-Quarter Cost Report every year, 12 months ahead out-of-sample forecasting is investigated for CCI in December starting from 1987 to 2008 based on their historical dataset respectively. For example, the subset dataset used to predict CCI in December 1987 contains historical data from January 1979 to December 1986. Holt ES and Holt-Winters ES models are fitted on each subset datasets, and then make 12 month ahead out-of-sample forecasting according to their best fitting models respectively.

Three general error measures Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), and Mean Absolute Error (MAE) are used to measure the predictability of ES models and the ENR's forecast. The less the error measures, the more accurate the prediction is. Table 5 summarizes the error measures for these two models. Figure 3 shows time series plot of actual CCI, ENR's forecast CCI, Holt ES forecast CCI, and Holt-Winters forecast CCI.

Table 5: Out-of-sample forecast error measures of ENR's Forecast & SARIMA

Error measures	ENR's Forecast	Holt ES	Holt-Winters ES
MAPE	1.22%	1.05%	1.07%
MSE	12910.00	10983.18	11149.05
MAE	73.33	68.64	69.59

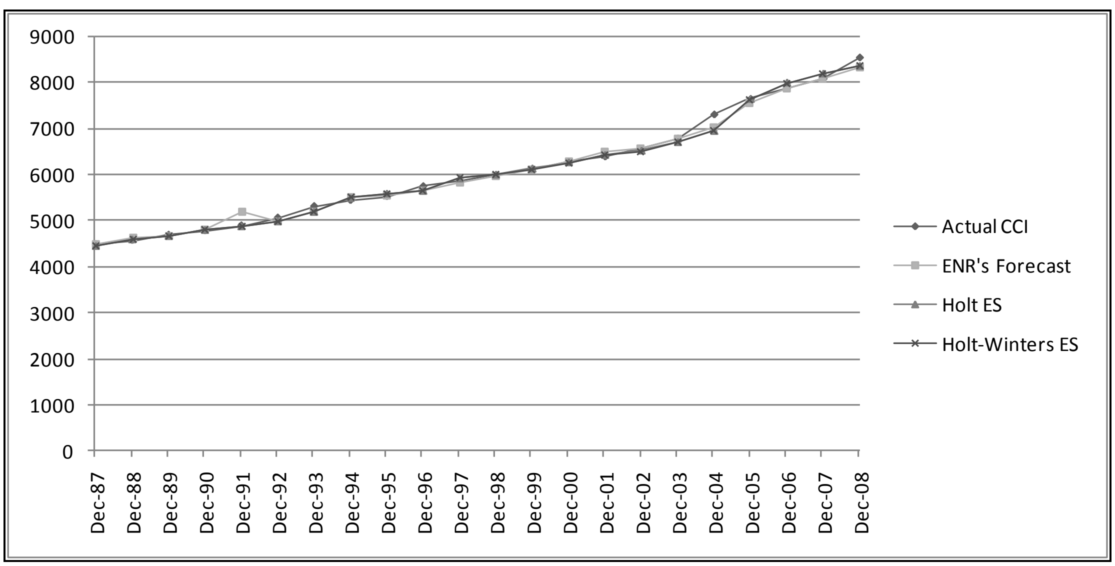

Figure 3. Comparison of forecast CCI with actual CCI value

It can be seen that both ENR's forecast, Holt ES and Holt-Winters forecast are close to the actual CCI values, presenting an accurate prediction with MAPE much smaller than 10%, which is an acceptable threshold recommended by Wong (2005). However, the predictability of both Holt ES model and Holt-Winters ES are even better than the ENR's forecast with smaller error measures.

Conclusions and Future Work

ENR CCI data shows increasing in the long-term and frequent variations in short-term. The ability to provide more accurate prediction of construction cost trends can lead to more-accurate bids and avoid under- or over-estimation. Our research in investigating applicability and predictability of exponential smoothing method on the historical CCI dataset shows that Holt ES and Holt-Winters ES model provides high level of accuracy in forecasting of CCI.

However, the predictability of Exponential Smoothing model does not perform well when CCIs make discrete jumps. For instance, Figure 3 shows both Holt ES and Holt-Winters ES underestimate the CCI in December 2004 and December 2008 when there is a large jump in CCI, which is an influential result of the general economic and financial crisis. Poisson (jump) processes (Dixit and Pindyck, 1994) should be tried in the future research to investigate whether they could create more realistic forecasting models to capture infrequent but discrete jumps of CCI.

With more accurate prediction of construction cost trend, contractors and owners can benefit in cost control and risk management by better-timed project execution. Exponential smoothing model can also be used by other researchers to develop predictive models for other construction indexes, such as the ENR's Building Cost Index.

Acknowledgements

This research is sponsored by an award from The Royal Institution of Chartered Surveyors (RICS) in The United Kingdom. The authors would like to thank Mr. Stephen Brown the Head of Research at RICS for his thoughtful comments. The authors would also like to thank Mr. Scott Lewis, the Editorial Research Director at *Engineering News-Record*, and Mr. Timothy Grogan, the Senior Editor of Construction Economics at *Engineering News-Record*, for their invaluable help in the process of data preparation for this research.

References

Akintoye, S. A., and Skitmore, R. M. (1994). "A comparative analysis of three macro price forecasting models." *Construction Management & Economics*, 12(3), 257-270.

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

Brockwell and Davis (2002), *Introduction to time series and forecasting*, Second Edition, Springer.

Chin, T. C., and Mital, D. P. (1998). "Time series modelling and forecasting of Singapore property price: an optimal control approach." Proceedings: 2nd International Conference on Knowledge-Based Intelligent Electronic Systems, IEEE Publishing, Adelaide, Australia, 370-375.

Dayton, K. J. (2006). "Recent Trends in Highway Construction Costs." 2006 WASHTO Subcommittee on Construction, Boise, Idaho.

Dixit, A., and Pindyck, R. (1994). *Investment under uncertainty*, Princeton University Press, Princeton, NJ.

Ervin, E. (2007). "How to protect profit as material prices rise." *Business Journal*

Fellows, R. F. (1991). "Escalation management: Forecasting the effects of inflation on building projects." *Construction Management and Economics*, 9(2), 187-204.

Gallagher, J. P. (2008). "Material Price Escalation: Managing the Risks." *Construction eNewsletter*.

Koehn, E., and Navvabi, M. H. (1989). "Economics and social factors in construction." *Cost Engineering*, 31(10), 15-18.

McCaffer, R., McCaffrey, M. J., and Thorpe, A. (1984). "Predicting the tender price of buildings in the early design stage: Method and validation." *The Journal of the Operational Research Society*, 35(5), 415-424.

Ng, S. T., Cheung, S. O., Skitmore, R. M., Lam, K. C., and Wong, L. Y. (2000). "The prediction of tender price index directional changes." *Construction Management and Economics*, 18(7), 843-852.

Runeson, K. G. (1988). "Methodology and method for price-level forecasting in the building industry." *Construction Management and Economics*, 6(1), 49-55.

Taylor, R. G., and Bowen, P. A. (1987). "Building price-level forecasting: an examination of techniques and applications." *Construction Management & Economics*, 5(1), 21-44.

Trost, S. M., and Oberlender, G. D. (2003). "Predicting accuracy of early cost estimates using factor analysis and multivariate regression." *ASCE Journal of Construction Engineering and Management*, 129(2), 198-204.

Tsay, R.S. (2005). *Analysis of financial time series*, Second Edition, John Wiley & Sons, Inc., Hoboken, NJ.

Wang, C.-H., and Mei, Y.-H. (1998). "Model for forecasting construction cost indices in Taiwan." *Construction Management and Economics* 16(2), 147-157.

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

Wong, M.W., Chan, P.C., and Chiang, Y.H (2005). "Time series forecasts of construction labor market in Hong Kong: the Box-Jenkins approach", *Construction Management and Economics* 23, 979-991.

Appendix A. Holt-Winters ES Formulation

Recursive equations for the model building procedure of the Holt-Winters ES approach are summarized below.

Level Smoothing:

$$L(t)=a\frac{Y(t)}{S(t-l)}+(1-a)[L(t-1)+T(t-1)]$$

Trend Smoothing:

$$T(t)=b[L(t)-L(t-1)]+(1-b)T(t-1)$$

Seasonal Smoothing

$$S(t)=c\frac{Y(t)}{L(t)}+(1-c)S(t-l)$$

Forecasting

$$\hat{Y}(t+1)=[L(t)+T(t)]S(t-l+1)$$

Where

- $t=1,2,3,\ldots,360$, is time index
- $Y(t)$,$L(t)$, $T(t)$, $\hat{Y}$ $(t+1)$ are the same as notations in the Holt ES model in section 4.2
- $S(t)$ is the seasonal index at time t
- l is the length of seasonality period, which is 12 for the CCI dataset
- $0 \leq a \leq 1$, $0 \leq b \leq 1$, and $0 \leq c \leq 1$ are constant weighting factors, which must be estimated such that the MSE of the error for this model is minimized

The initial values for trend smoothing, level smoothing, and forecasting are:

The initial value of the trend factor:

$$T(1)=\frac{1}{l}\left(\frac{Y(l+1)-Y(1)}{l}+\frac{Y(l+2)-Y(2)}{l}+\cdots+\frac{Y(2l)-Y(l)}{l}\right)$$

The initial value of the level factor:

$$L(1)=\frac{1}{l}\sum_{i=1}^{l}Y(i)-\frac{l}{2}T(1)$$

The initial values of the seasonal factor:

$$S(t)=\frac{1}{m}\left(\frac{Y(t)}{\sum_{i=1}^{l}Y(i)}+\frac{Y(t+l)}{\sum_{i=l+1}^{2l}Y(i)}+\cdots+\frac{Y(t+(m-1)l)}{\sum_{i=(m-1)l+1}^{ml}Y(i)}\right),\ t=1,2,\ldots,l$$

m is number of available seasonality in dataset, for CCI dataset m=23