ANALYSIS OF CONSTRUCTION COST VARIATIONS USING MACROECONOMIC, ENERGY AND CONSTRUCTION MARKET VARIABLES

A Dissertation
Presented to
The Academic Faculty

by

Seyed Mohsen Shahandashti

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Building Construction

Georgia Institute of Technology
August 2014

Copyright © 2014 by Seyed Mohsen Shahandashti

ANALYSIS OF CONSTRUCTION COST VARIATIONS USING MACROECONOMIC, ENERGY AND CONSTRUCTION MARKET VARIABLES

Approved by:

Dr. Baabak Ashuri, Advisor
School of Building Construction
Georgia Institute of Technology

Dr. Daniel Castro-Lacouture
School of Building Construction
Georgia Institute of Technology

Prof. Kathy O. Roper
School of Building Construction
Georgia Institute of Technology

Dr. David Goldsman
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Paul Goodrum
School of Civil, Environmental and Architectural Engineering
University of Colorado at Boulder

Date Approved: June 08, 2014

Dedicated to
My Mother Akram and My Father Hossein

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Baabak Ashuri, for his continuous encouragement, immense knowledge, invaluable support, endless patience and excessive interest which made me energetic to continue studying and working hard.

I would like to thank the members of my PhD examination committee, Dr. Daniel Castro-Lacouture, Professor Kathy Roper, Dr. Paul Goodrum, and Dr. David Goldsman for their encouragement, insightful comments, and great questions. I would also like to thank Prof. Godfried Augenbroe for his insightful comments about my research activities at Georgia Tech.

I would like to thank my friends and colleagues at Georgia Tech, Hamed Kashani, Jian Lu, Kia Mostaan, Mohammad Illbeigi, Saman Yarmohammadi, Mostafa Reisi, Atefe Makhmalbaf, Anna Losonczi, Miklos Oroszlany, and Yeonsook Heo for all the great discussions and fun we had.

Last but not least, I am grateful to my parents, Akram Razaghdoost and Hossein Shahandashti who are the most important and dearest people in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ADF	Augmented Dickey-Fuller
AHE	Average Hourly Earnings
AIC	Akaike Information Criterion
ARCH	Autoregressive Conditional Heteroskedasticity
ARIMA	Autoregressive Integrated Moving Average
BP	Building Permits
CCI	Construction Cost Index
CPI	Consumer Price Index
COP	Crude Oil Price
COP	Construction Spending
DJIA	Dow Jones Industrial Average
ELC	Employment Level in Construction
ENR	Engineering News Record
FFR	Federal Funds Rate
FHWA	Federal Highway Administration
GBM	Geometric Brownian Motion
GDP	Gross Domestic Product
GDPIPD	Gross Domestic Product Implicit Price Deflator
HS	Housing Starts
HW-ES	Holt-Winters Exponential Smoothing
HUD	Housing and Urban Development
MAPE	Mean Absolute Prediction Error
MS	Money Supply

MSE	Mean Squared Error
NHCCI	National Highway Construction Cost Index
PLR	Prime Loan Rate
PPI	Producer Price Index
UR	Unemployment Rate
VEC	Vector Error Correction
VECM	Vector Error Correction Model
TPI	Tender Price Index

SUMMARY

Recently, construction cost variations have been larger and less predictable. These variations are apparent in trends of indices such as Engineering News Record (ENR) Construction Cost Index (CCI) and National Highway Construction Cost Index (NHCCI). These variations are problematic for cost estimation, bid preparation and investment planning. Inaccurate cost estimation can result in bid loss or profit loss for contractors and hidden price contingencies, delayed or cancelled projects, inconsistency in budgets and unsteady flow of projects for owner organizations. Cost variation has become a major concern in all industry sectors, such as infrastructure, heavy industrial, light industrial, and building. The major problem is that construction cost is subject to significant variations that are difficult to forecast. The objectives of this dissertation are to identify the leading indicators of CCI and NHCCI from existing macroeconomic, energy and construction market variables and create appropriate models to use the information in past values of CCI and NHCCI and their leading indicators in order to forecast CCI and NHCCI more accurately than existing CCI and NHCCI forecasting models.

A statistical approach based on multivariate time series analysis is used as the main research approach. The first step is to identify leading indicators of construction cost variations. A pool of 16 candidate (potential) leading indicators is initially selected based on a comprehensive literature review about construction cost variations. Then, the leading indicators of CCI are identified from the pool of candidate leading indicators using empirical tests including correlation tests, unit root tests, and Granger causality tests. The identified leading indicators represent the macroeconomic and construction

market context in which the construction cost is changing. Based on the results of statistical tests, several multivariate time series models are created and compared with existing models for forecasting CCI. These models take advantage of contextual information about macroeconomic condition, energy price and construction market for forecasting CCI accurately. These multivariate time series models are rigorously diagnosed using statistical tests including Breusch–Godfrey serial correlation Lagrange multiplier tests and Autoregressive conditional heteroskedasticity (ARCH) tests. They are also compared with each other and other existing models. Comparison is based on two typical error measures: out-of-sample mean absolute prediction error and out-of-sample mean squared error.

Based on the unit root tests and Granger causality tests, consumer price index, crude oil price, producer price index, housing starts and building permits are selected as leading indicators of CCI. In other words, past values of these variables contain information that is useful for forecasting CCI. Based on the results of cointegration tests, Vector Error Correction (VEC) models are created as proper multivariate time series models to forecast CCI. Our results show that the multivariate time series model including CCI and crude oil price pass diagnostic tests successfully. It is also more accurate than existing models for forecasting CCI in terms of out-of-sample mean absolute prediction error and out-of-sample mean square error.

The predictability of the multivariate time series modeling for forecasting CCI is also evaluated using stochastically simulated data (Simulated CCI and crude oil price). First, 50 paths of crude oil price are created using Geometric Brownian Motion (GBM). Then, 50 paths of CCI are created using Gaussian Process that is considering the relationship between CCI and crude oil price over time. Finally, 50 multivariate and univariate time series models are created using the simulated data and the predictability of univariate and multivariate time series models are compared. The results show that the multivariate modeling is more accurate than univariate modeling for forecasting simulated CCI. The sensitivity of the models to inputs is also examined by adding errors to the simulated data and conducting sensitivity analysis.

The proposed approach is also implemented for identifying the leading indicators of NHCCI from the pool of candidate leading indicators and creating appropriate multivariate forecasting models that use the information in past values of NHCCI and its leading indicators. Based on the unit root tests and Granger causality tests, crude oil price and average hourly earnings in the construction industry are selected as leading indicators of NHCCI. In other words, past values of these variables contain information that is useful for forecasting NHCCI. Based on the results of cointegration tests, Vector Error Correction (VEC) models are created as the proper multivariate time series models to forecast NHCCI. The results show that the VEC model including NHCCI and crude oil price, and the VEC model including NHCCI, crude oil price, and average hourly earnings pass diagnostic tests. These VEC models are also more accurate than the univariate models for forecasting NHCCI in terms of out-of-sample prediction error and out-of-sample mean square error.

The findings of this dissertation contribute to the body of knowledge in construction cost forecasting by rigorous identification of the leading indicators of construction cost variations and creation of multivariate time series models that are more accurate than the existing models for forecasting construction cost variations. It is expected that proposed forecasting models enhance the theory and practice of construction cost forecasting and help cost engineers and capital planners prepare more accurate bids, cost estimates and budgets for capital projects.

CHAPTER 1

INTRODUCTION

Research Background

Construction costs are subject to variations during various phases of construction (Figure 1.1). The expected increases in construction cost should be considered in the project budget and utilized when preparing budget submissions for funding approval.

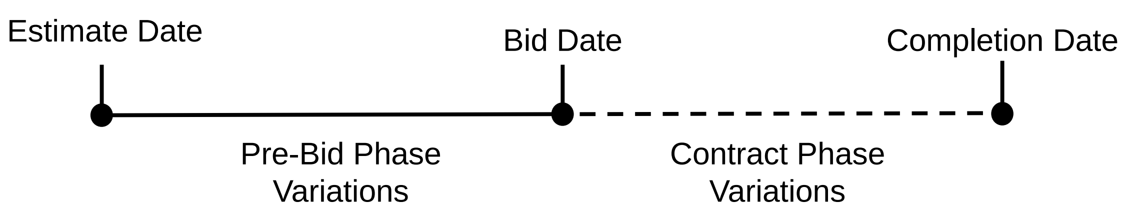


Figure 1.1: Phases of Construction Cost Variations

Construction cost variations have caused several practical challenges for owners and contractors. Around two-third of contractors believe that variability in construction costs is one of the most important risks that impact their profits (Ervin 2007). Construction cost variations have adverse impact on public and private owners (Dayton 2006; Gallagher 2008; O'Haren 2004). "Fluctuations in prices for items such as steel, Portland cement, asphalt, and fuel have been larger and less predictable than was typical in the past. State transportation budgets have been inconsistent from year to year, resulting in an unsteady flow of projects" (FHWA 2012). Moreover, increase in frequency of large-scale buildings becomes a concern due to lengthy construction periods and presence of recent volatile fluctuations of construction material prices (Hwang et al. 2012).

Cost variations have become a major concern in all construction industry sectors, such as infrastructure, heavy industrial, light industrial, and building. Construction cost variations are problematic for cost estimation, bid preparation, and investment planning. Inaccurate cost estimation can result in bid loss or profit loss for contractors. It can also result in hidden price contingencies, delayed or cancelled projects, inconsistency in budgets, and unsteady flow of projects for owner organizations. The major problem is that construction cost is subject to significant variations that are difficult to forecast.

Construction cost indices have been used to measure the cost trends in the construction industry (Wilmot and Cheng 2003; Touran and Lopez 2006). Construction Cost Index (CCI) is one of the oldest, most important and commonly used construction cost indices in the United States (Touran and Lopez 2006). CCI, which has been published by Engineering News-Record (ENR) monthly in the United States, is defined as the weighted aggregate of average prices of constant quantities of common labour, standard structural steel, Portland cement and lumber in 20 cities (ENR 2014). CCI is widely used for cost estimation and budgeting of capital projects in the United States. ENR just published a report celebrating 100 years of ENR cost indexes (ENR 2013). The ENR report highlights several testimonials showing the wide adoption of ENR construction cost indexes. For example, Jerry Welch, chief of the Cost Relocations Team, U.S. Army Engineer District, Memphis stated that the "ENR indexes are an excellent resource" and expressed that he used ENR construction cost indexes "to accurately escalate contract pricing for the future where the Corps Index is not really applicable" (ENR 2013). CCI is subject to short- and medium-term variations that are problematic for cost estimation and bid preparation. Figure 1.2 presents the CCI variations in 2002. The high level question is how to forecast variations in CCI accurately.

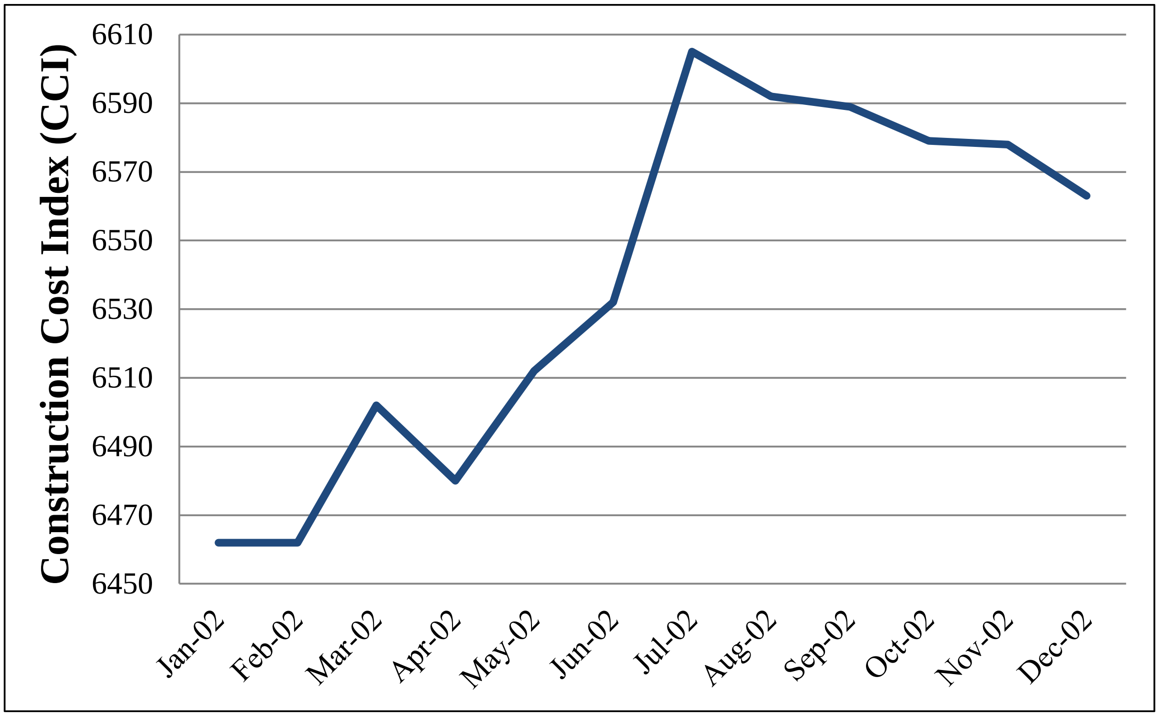


Figure 1.2: CCI variations in 2002

National Highway Construction Cost Index (NHCCI) is another widely used construction cost index. Federal Highway Administration (FHWA) has published National Highway Construction Cost Index (NHCCI) quarterly since 2003. NHCCI measures the average change over time in the roadway construction material and service prices paid by State transportation departments (FHWA 2013). NHCCI is also subject to significant variations that make forecasting problematic.

State of Knowledge in Forecasting Construction Cost Variations

Quantitative methods have been proposed to forecast construction cost indices, such as CCI and NHCCI. These methods can be classified into two major categories (Touran and Lopez 2006): Causal methods and statistical methods.

Statistical Methods

The proposed statistical methods use time series analysis and curve fitting to forecast construction cost indices (Hanna and Blair 1993). For example, Ashuri and Lu (2010) compared various univariate time series models to forecast CCI. They concluded that seasonal autoregressive integrated moving average model and Holt-Winters exponential smoothing are the most accurate univariate time series approaches for in-sample and out-of-sample forecasting of CCI, respectively. The statistical methods do not have explanatory capability and they are just suitable for short-term forecasting (Touran and Lopez 2006; Goh and Teo 2000).

Causal Methods

The causal methods forecast construction cost indices based on their explanatory relationship with other variables. For example, in a rare study for forecasting CCI using multiple 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 changes in CCI. He concluded that CCI prediction is a complex problem and CCI cannot be accurately predicted by the neural network models. The comprehensive literature review about the causal and statistical methods for forecasting construction cost indices are provided in Chapter 2 (Literature Review).

Gaps in Knowledge

Statistical Methods

Current statistical methods, such as univariate time series models, do not have explanatory capability and they are just suitable for short-term forecasting (Wong and Ng 2010; Touran and Lopez 2006; Goh and Teo 2000; Pindyck and Rubinfeld 1998).

Causal Methods

The forecasting power of causal models depends on the identification of appropriate leading variables (Wong and Ng 2010; Williams 1994). However, the proven leading variables of construction cost indices are not known. Moreover, the temporal relations of variables are ignored in these models.

Leading Indicator

A leading indicator is a time series that is highly correlated with another time series when it is lagged (Akintoye et al. 1998). In other words, the leading indicator reaches its cyclical turning points earlier than cyclical turning points of the other time series. With respect to the Granger causality (Granger 1969), a leading indicator of a variable can be defined as an indicator which past values contain information that is useful for forecasting the future values of the variable.

Lagging Indicator

On the other hand, a lagging indicator is a time series that is highly correlated with another time series when the other time series is lagged. In other words, the lagging indicator reaches its cyclical turning points later than cyclical turning points of the other time series.

Points of Departure

This research departs from the literature by identifying the leading indicators of CCI and NHCCI and creating multivariate time series models that use information in multiple time series (CCI, NHCCI and their leading indicators) to forecast CCI and NHCCI more accurately than the existing univariate time series models. These multivariate models take into account the temporal relationship between construction cost indices (i.e., CCI and NHCCI) and their proven leading indicators in order to forecast construction cost indices (i.e., CCI and NHCCI) accurately. In addition, this research departs from the literature by using simulation to test predictability of models to forecast CCI.

Research Objectives

The objectives of this research are:

(1) Identify the leading indicators of CCI and NHCCI from existing macroeconomic, energy and market variables.

(2) Create appropriate models to use the information in past values of construction cost indices (i.e., CCI and NHCCI) and their leading indicators in order to forecast the construction cost indices (i.e., CCI and NHCCI) more accurately than existing CCI and NHCCI forecasting models.

Research Approach

A statistical approach based on multivariate time series analysis is used as the main research approach. The following steps are taken to achieve the objectives of this research:

- Identify the leading indicators of CCI.
    - Select candidate (potential) leading indicators using literature regarding the explanatory variables of construction cost and price indices that are used around the world.
    - Use statistical tests to identify leading indicators of CCI from the pool of the potential leading indicators.
- Create appropriate multivariate time series models to forecast CCI
    - Select an appropriate type of multivariate model to forecast CCI
    - Create various multivariate time series models using CCI and various combinations of the identified leading indicators.
    - Diagnose the created models
    - Validate the results using testing data

- Test predictability of the multivariate time series modeling for forecasting CCI using simulated data (Simulated CCI and leading indicators).
- Implement and validate the proposed approach for creating appropriate multivariate time series models for forecasting National Highway Construction cost Index (NHCCI).

Organization of Dissertation

Chapter 2 provides the comprehensive review of literature. This chapter reviews literature regarding CCI forecasting, NHCCI forecasting, and explanatory variables of construction cost and price indices. Sixteen candidate (potential) leading indicators of construction cost variations are identified in this section.

Chapter 3 shows how the leading indicators of CCI are identified from the pool of candidate leading indicators using statistical tests including correlation tests, unit root tests, and Granger causality tests. The identified leading indicators represent the macroeconomic and construction market context in which the construction cost is changing.

Chapter 4 shows how the appropriate type of multivariate time series models is selected based on statistical tests. Several multivariate time series models are also created in Chapter 4. These models are validated through comparison with the best existing univariate models for forecasting CCI.

The predictability of the multivariate time series modeling for forecasting CCI is tested using stochastically simulated data in Chapter 5. The purpose of Chapter 5 is to simulate CCI and its relevant leading indicators and show that multivariate time series modeling provides better solution than univariate time series modeling for forecasting CCI.

The leading indicators of NHCCI are selected using statistical tests in Chapter 6. This chapter also shows how appropriate type of multivariate time series models for forecasting NHCCI is selected based on the statistical tests. Several multivariate time series models for forecasting NHCCI are created in this chapter. These models are validated through comparison with univariate time series models for forecasting CCI.

The findings of this dissertation are thoroughly discussed in Chapter 7. More specifically, the impacts of crude oil price, macroeconomic condition and construction market on construction cost variations are discussed. Some explanations are also provided to describe the results of statistical tests.

Conclusions are presented in Chapter 8. The contributions of this research to the state of knowledge and the state of practice are explicitly expressed. Recommendations are also provided for further research.

CHAPTER 2

LITERATURE REVIEW

This chapter reviews literature regarding CCI forecasting, NHCCI forecasting, and explanatory variables of construction cost and price indices. Sixteen candidate (potential) leading indicators of construction cost variations are identified and presented at the end of this chapter.

Forecasting Construction Cost Index

Quantitative methods have been proposed to forecast construction cost indices. These methods can be classified into two major categories (Touran and Lopez 2006): Causal and statistical methods. The proposed statistical methods use time series analysis and curve fitting to forecast construction cost indices (Hanna and Blair 1993). Ashuri and Lu (2010) compared various univariate time series models to forecast CCI. They concluded that seasonal autoregressive integrated moving average model and Holt-Winters exponential smoothing are the most accurate univariate time series approaches for in-sample and out-of-sample forecasting of CCI, respectively. The statistical methods 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 indices based on their explanatory relationship with other variables. In a rare study for forecasting CCI using multiple 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 changes in CCI. He concluded that CCI prediction is a complex problem and CCI cannot be accurately predicted by the neural network models.

Forecasting National Highway Construction Cost Index

Pewdum et al (2009) developed neural network models to forecast final budget and duration of a highway construction project during construction stage. Williams (2005) developed regression and neural network models to predict the cost of highway projects with bid ratios as input variables. Wilmot and Mei (2005) created an artificial neural network model that relates overall highway construction costs to the cost of construction material, labor, and equipment, the characteristics of the contract and the contracting environment. Al-Tabtabai et al. (1999) developed neural networks to predict preliminary cost estimation of highway construction using the following input variables: location, utilities, soil nature, type of consultant, detour construction, hauling distance, financial condition, type of road, and need. Hegazy and Ayed (1998) also developed a neural network model to estimate cost of highway projects. The model inputs are description of project size, year of construction, project location, capacity, and other uncertainty-related factors.

Wilmot and Cheng (2003) developed a regression model to predict overall highway construction costs after analyzing the Louisiana Highway Construction Index. Chengalur-Smith et al (1997) created a non-linear regression model to estimate bridge rehabilitation. The model had eight explanatory variables including region, bridge type, deck area, substructure area, age, functional class, component condition index, and completed work. Sthapit and Mori (1994) created a parametric model to estimate highway earthwork cost that depends on cutting depth, cross slope, and soil type in the construction area. Sanders et al. (1992) have also created simple regression models for work items that are greater than 1% of the total project cost. Saito et al (1991) also developed regression models for the estimation of bridge replacement costs where bridge attributes were the independent variables. Bell (1987) created multiple linear regression models for estimating preliminary cost using a list of predictors representing line items.

Accurate forecasting of NHCCI requires knowledge about historical variations of highway construction costs and temporal relationships of these variations with fluctuations in some explanatory variables representing macroeconomic and construction market conditions. However, existing literature lacks rigorous identification of explanatory variables that are useful to predict NHCCI. Explanatory variables that are useful to forecast NHCCI are the leading indicators of NHCCI. These explanatory variables are called leading indicators because they contain information that is useful to predict future values of NHCCI. These leading indicators should be included in appropriate multivariate models for forecasting NHCCI. These multivariate models should take into account the temporal relationship between NHCCI and the proven leading indicators in order to forecast NHCCI accurately. Existing models do not take into account both proven leading indicators of NHCCI and the temporal relationship among variables.

Gaps in Knowledge for Forecasting Construction Cost Variations

Statistical Methods

Current statistical methods, such as univariate time series models, do not have explanatory capability and they are just suitable for short-term forecasting (Wong and Ng 2010; Touran and Lopez 2006; Goh and Teo 2000; Pindyck and Rubinfeld 1998).

Causal Methods

The forecasting power of causal models depends on the identification of appropriate leading variables (Wong and Ng 2010; Williams 1994). However, the proven leading variables of construction cost indices are not known. Moreover, the temporal relations of variables are ignored in these models.

Explanatory Variables of Construction Cost and Price Indices

Explanatory variables of construction cost and price indices play an important role in forecasting. This section is devoted to review the literature regarding the explanatory variables of construction cost and price indices. After reviewing the literature, a list of potential (candidate) leading indicators of construction cost is developed. Identification of this pool of candidate leading indicators from the literature is the first steps towards creating appropriate multivariate construction cost forecasting models.

Research regarding the explanatory variables of CCI is limited. In a rare study, Williams (1994) used the trends in CCI, prime lending rate, housing starts, and the months of the year to predict changes in CCI.

Despite limited research for finding the leading indicators of CCI, several studies were conducted to determine the leading indicators of building price indices, such as Tender Price Index (TPI). TPI is an output index representing the average price that clients and/or owners need to pay to build a facility (Ng et al., 2000 and Wong and Ng, 2010). TPI is published in the United Kingdom, Hong Kong, and Singapore and used by clients as an indication of construction cost level (Rowlinson and Walker, 1994). The leading indicators of TPI have potential to be among the leading indicators of CCI and NHCCI. Therefore, the leading indicators of TPI are reviewed and included in the pool of potential leading indicators of construction cost variations.

Taylor and Bowen (1987) analyzed the Index of Building Cost, a tender price index provided by Bureau of Economic Research (BER). They emphasized the importance of demand-based factors for formulating future building price indices. They concluded that construction supply capability has a long-term effect on the price movements. Skitmore (1987) also found a positive relationship between the new orders representing the construction demand and price level.

Runeson (1988) determined that movements in building prices are the product of changes in input prices and changes in prices driven by market conditions. He represented market conditions as a function of three independent variables: building approvals, fixed capital formation in buildings, and unemployment rate. Fellows (1988) used studied the leading indicators of construction price in the United Kingdom. He concluded that interest rates, investment intentions, architect's new commissions, production drawings, enquiries, orders, expected volume of work, and building cost are the leading indicators of the construction price in the U.K.

Akintoye et al. (1998) summarized unemployment level, construction output, industrial production, and the ratio of price to cost indices in manufacturing as the consistent leading indicators of TPI. Ng et al. (2000) used the pattern of changes in eight leading indicators (best lending rate, building cost index, consumer price index, gross domestic product, gross domestic product of construction industry, implicit gross domestic product deflation, money supply, and employment rate) for predicting the direction of changes in the Hong Kong TPI.

In a recent study, Wong and NG (2010) selected eight potential indicators (bank interest rate, building cost index, composite consumer price index, gross domestic product, gross domestic product in construction, implicit gross domestic product deflator, money supply, and unemployment rate) for identifying explanatory variables of TPI. They concluded that building cost index, gross domestic product, and gross domestic product in construction have explanatory value for predicting TPI.

Table 2.1 summarizes sixteen commonly used explanatory variables of the construction cost, their brief descriptions, and the sources from which the data can be retrieved in the U.S. These variables represent the U.S. construction market and macroeconomic conditions and might be useful to predict CCI and NHCCI. Our hypothesis is that these widely used and publicly available variables can be used along with CCI and NHCCI within appropriate multivariate time series models to forecast CCI and NHCCI more accurately than the existing univariate time series models.

Table 2.1: Potential leading indicators for construction cost variations, respective IDs, and data sources

Candidate leading Indicators	**ID**	**Source**
Consumer Price Index	CPI	U.S. Bureau of Labor Statistics
Federal Funds Rate	FFR	Board of Governors of the Federal Reserve Systems
Unemployment Rate	UR	U.S. Bureau of Labor Statistics
Employment Level in Construction	ELC	U.S. Bureau of Labor Statistics
Average Weekly Hours	AWH	U.S. Bureau of Labor Statistics
Prime Loan Rate	PLR	Board of Governors of the Federal Reserve Systems
Building Permits	BP	U.S. Bureau of Census
Money Supply	MS	Board of Governors of the Federal Reserve Systems
Average Hourly Earnings	AHE	U.S. Bureau of Labor Statistics
Dow Jones Industrial Average	DJIA	Yahoo Finance
Crude Oil Price	COP	U.S. Energy Information Administration
Producer Price Index	PPI	U.S. Bureau of Labor Statistics
Housing Starts	HS	U.S. Bureau of Census
Construction Spending	CS	U.S. Census Bureau
Gross Domestic Product	GDP	U.S. Bureau of Economic Analysis
GDP Implicit Price Deflator	GDPIPD	U.S. Bureau of Economic Analysis

The sixteen indicators shown in Table 2.1 are selected as prime candidates for leading indicators of construction cost variations:

- Prime loan rate
- Housing starts (number of housing starts)
- Building permits (number of building permits)
- Unemployment rate
- Consumer price index
- Producer price index
- Gross domestic product
- GDP implicit price deflator
- Money supply
- Construction spending
- Federal funds rate
- Dow Jones industrial average
- Crude oil price
- Employment level in construction
- Average hourly earnings, and
- Average weekly hours

The above variables depict the national macroeconomic conditions, energy cost, and construction market conditions.

Indicators Representing U.S. Macroeconomic Condition

The following candidate leading indicators represent U.S. macroeconomic conditions from various perspectives:

- Prime loan rate

- Unemployment rate
- Consumer price index
- Producer price index
- Gross domestic product
- GDP implicit price deflator
- Money supply
- Federal funds rate

While GDP represents national income and economic health of the U.S., consumer price index and GDP implicit price deflator are widely used to represent inflation at the national level.

Money supply is a type of measure that represents the amount of money in the nation’s economy. In the U.S., at least two types of money supply’s measures are tracked: M1 and M2. M1 consists of currency in the public banks, institutions and the U.S treasury, traveler’s checks, demand deposits, and other checkable deposits. M2 consists of M1 as well as savings deposits, time deposits less than $100, and balances in retail money market mutual funds. We used M2, which is a broader measure. Factors representing interest rates at the national level are among important macroeconomic indicators. Prime loan rate and federal funds rate are two widely used measures representing interest rates.

Indicators Representing U.S. Construction Market

The following candidate leading indicators represent construction market conditions in the U.S. from various perspectives:

- Housing starts
- Building permits
- Construction spending

- Employment level in construction,
- Average hourly earnings, and
- Average weekly hours

Number of new privately owned housing that their construction has been authorized (housing permits) or started (housing starts) provides useful information about expected construction activity in a relatively near future.

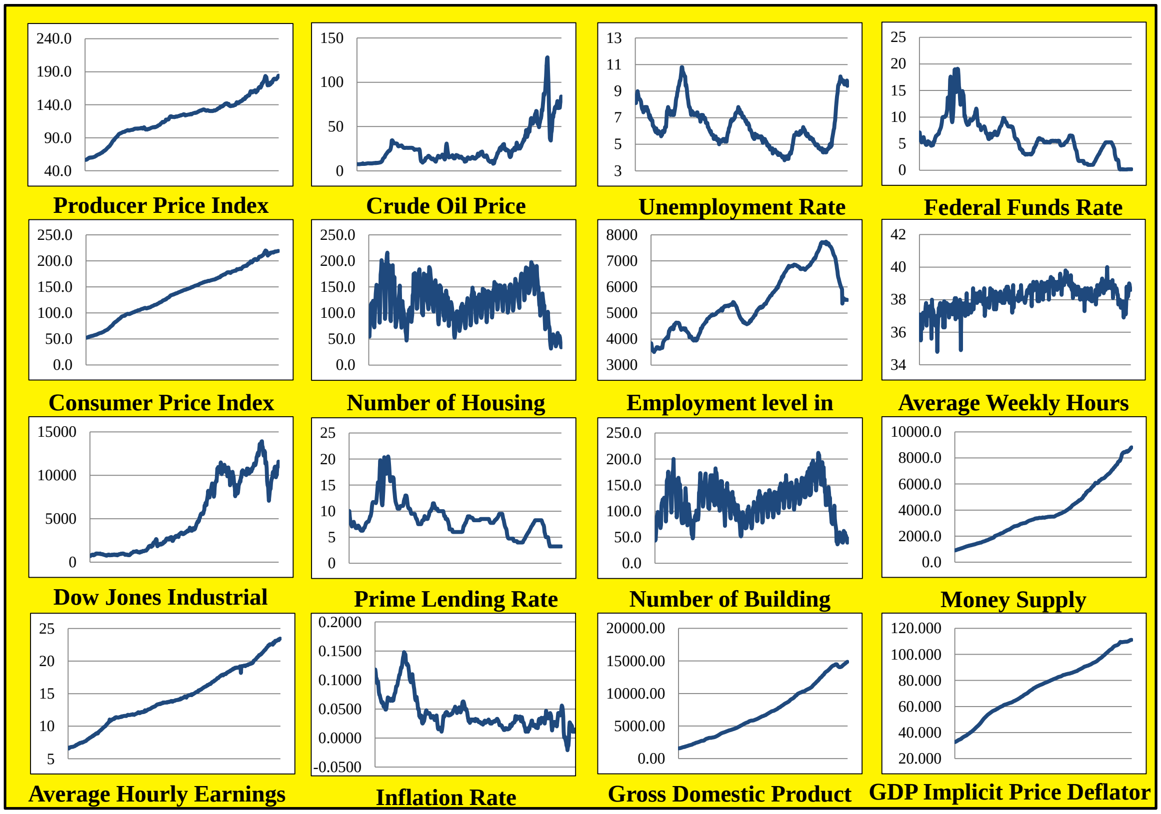
Construction spending is a measure of the value of new construction activities including residential projects, non-residential projects, and public projects. Unemployment rate represent the U.S. labor force at the national level. Employment level in construction is a useful measure to represent labor force just in the construction sector of the economy.

Indicators Representing U.S. Energy Market

Energy price level is widely ignored as one of the potential leading indicators of construction cost. Crude oil price is used as a measure for representing energy price level. Assessment of stock market indices as the leading indicators of construction cost is another interesting exploration since they are widely available and popular. Among stock market indices, Dow Jones Industrial Average is widely quoted stock market index (Tse, 1999) and assessed as one of the candidate leading indicators.

Figure 2.1 presents the trends of the potential leading indicators of construction costs. As it is clear in this figure, the potential leading indicators that represent macroeconomic, energy and construction market variable have various trends.

Figure 2.1: Trends of the potential leading indicators of construction cost



CHAPTER 3

IDENTIFYING LEADING INDICATORS OF CONSTRUCTION COST INDEX

An important step towards forecasting the CCI trends is to identify its leading indicators. The objective of this chapter is to identify the leading indicators of CCI. Empirical tests are used to identify the leading indicators of CCI from the pool of candidate (potential) leading indicators representing the U.S. construction and economic environment. Initially, the relevance of the candidate leading indicators is tested using Pearson Correlation Analysis (Wong and Ng, 2010). Secondly, multivariate time series tests are conducted to identify the leading indicators of CCI from the pool of candidate leading indicators. All the potential leading indicators are time series. Time series tests are usually 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 for identifying the order of integration of the potential leading indicators. The order of integration of the variables plays an important role in the implementation of multivariate time series tests.

Granger causailty test is used to identify the leading indicator of CCI from the pool of candidate (potential) leading indicators. Researchers use Granger causality test to study the lead-lag relationship between economic and construction variables. For instance, Granger causality test was used to examine the effects of fluctuations in the money supply on the fluctuations in construction activity flows in Hong Kong (TSE and Raftery, 2001). It was used to evaluate the causal relationship between construction and other economic sectors in Singapore (Lean, 2001). It was applied to study the effects of shocks in construction outputs on major economic indicators in Singapore (Chan, 2002). Wong et al. (2008) used Granger causality test to show how construction outputs (measured by gross value of construction works) drive the economic growth (measured by Gross Domestic Product (GDP)) in Hong Kong. Wong and Ng (2010) used Granger causality test to identify the leading indicators of tender price index in Hong Kong (2010). Granger causality test has not been applied for identifying the leading indicators of CCI.

Construction Cost Indices – collected monthly by Engineering News Record from January 1975 to December 2008 – are used as CCI time series data in our empirical study. CCI is defined based on the 20-city average prices of 200 hours of common labor, 25 cwt of fabricated standard structural steel, 1.128 tons of bulk Portland cement, and 1.088 board-ft. of 2×4 lumber (ENR, 2014). A total of 408 data points that represent the values of CCI in every month from January 1975 to December 2008 are used for identifying the leading indicators of CCI.

R is used as the statistical software for conducting the time series tests and creating the time series models. R is widely used for implementing time series analysis. It is popular among statisticians in the academia and business. It is freely available and well documented, which makes it interesting for statistical analysis. R provides free access to the source code. This feature provides the opportunity to make sure that the theory and software match well and the outcomes are meaningful.

Correlation Analysis

The Pearson correlation analysis is used as the preliminary analysis tool to test the relevance of the above potential leading indicators to CCI. Table 3.1 summarizes the results of the correlation tests. Based on the correlation results and test statistics, consumer price index (+0.99), federal funds rate (-0.59), unemployment rate (-0.63), employment level in construction (+0.94), Average work hours (+0.71), prime loan rate (-0.49), money supply (+0.98), average hourly earnings (+0.99), Dow Jones industrial average (+0.92), crude oil price (+0.66), producer price index (+0.99), building permits (0.24), construction spending (+0.97), GDP (+0.99), and GDP implicit price deflator (+0.99) are significantly correlated with CCI at 1% significance level. These results support the selection of the above variables as prime candidates for leading indicators. Number of housing starts (+0.04) is the only variable that is not significantly correlated with CCI. However, it is a variable that has been considered significant in the literature (Taylor and Bowen, 1987; Skitmore, 1987; and Runeson, 1988). Therefore, it is decided to keep it as a potential leading indicator for further analysis.

Table 3.1: Results of Pearson correlation tests between CCI and potential leading indicators of CCI

Variables	Corr.	Test Statistic
CCI,CPI	0.99	167.46**
CCI,FFR	-0.59	-14.88**
CCI,UR	-0.63	16.27**
CCI,ELC	0.94	57.32**
CCI,AWH	0.71	20.39**
CCI,PLR	-0.49	-11.33**
CCI,BP	0.24	4.88**
CCI,MS	0.98	109.66**
CCI,AHE	0.99	214.90**
CCI,DJIA	0.92	47.60**
CCI,COP	0.66	17.92**
CCI,PPI	0.99	120.17**
CCI,HS	0.04	0.76
CCI,CS	0.97	83.54**
CCI,GDP	0.99	127.58**
CCI,GDPIPD	0.99	149.07**
CCI,AHE	0.99	214.90**

*Notes: * indicates rejection of the null hypothesis at the 5% significance level; ** indicates rejection of the null hypothesis at the 1% significance level*

Figures 3.1 to 3.15 show the scatter plots representing the correlation relationships between CCI and the candidate leading indicators.

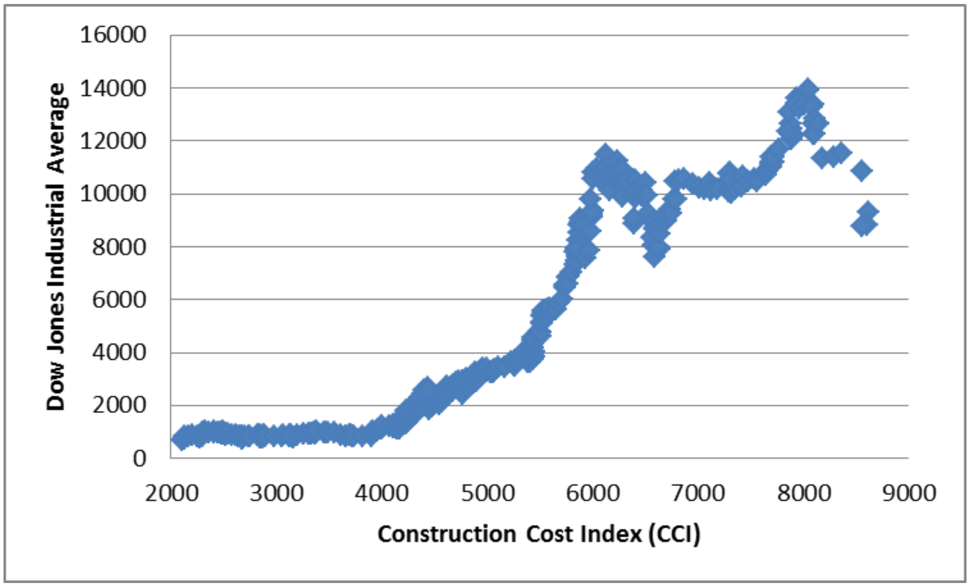


Figure 3.1: Scatter plot representing the relationship between CCI and Dow Jones Industrial Average

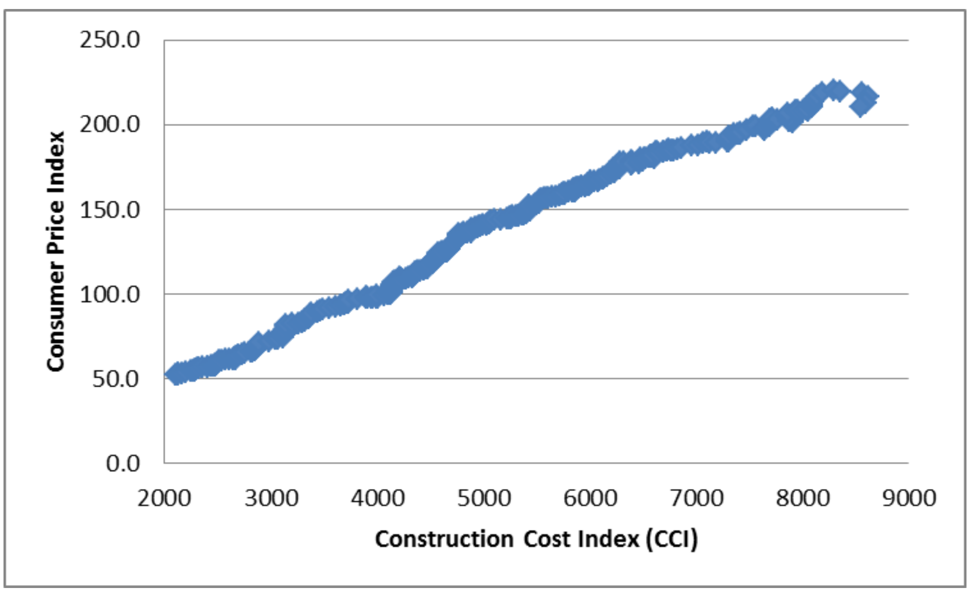


Figure 3.2: Scatter plot representing the relationship between CCI and Consumer Price Index

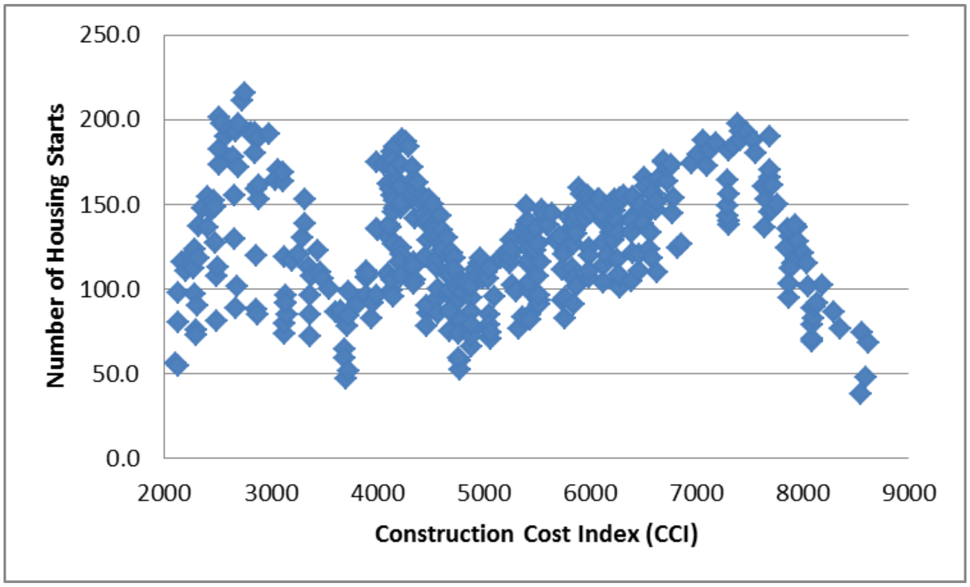


Figure 3.3: Scatter plot representing the relationship between CCI and number of housing starts [Thousands of units]

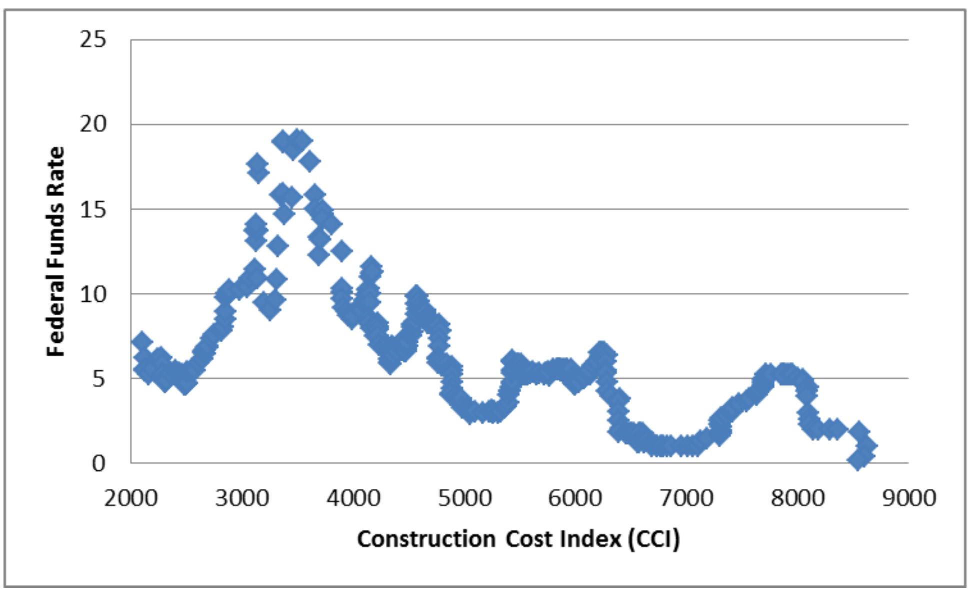


Figure 3.4: Scatter plot representing the relationship between CCI and federal funds rate

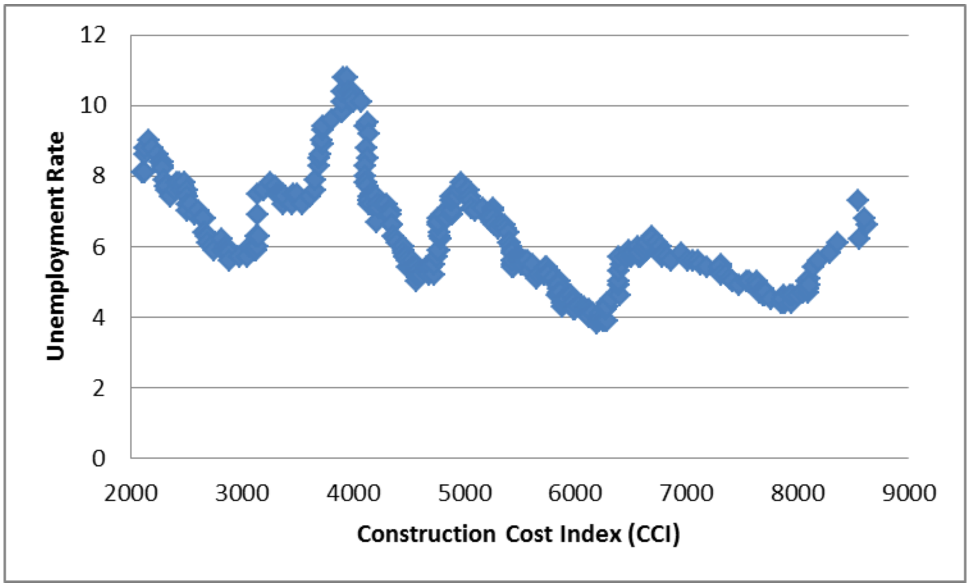


Figure 3.5: Scatter plot representing the relationship between CCI and unemployment rate

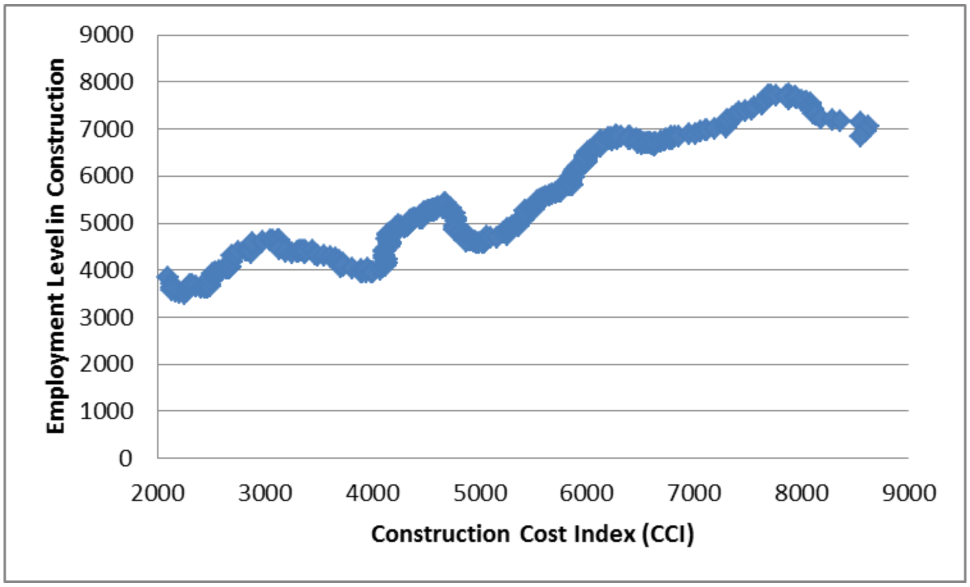


Figure 3.6: Scatter plot representing the relationship between CCI and employment level in construction

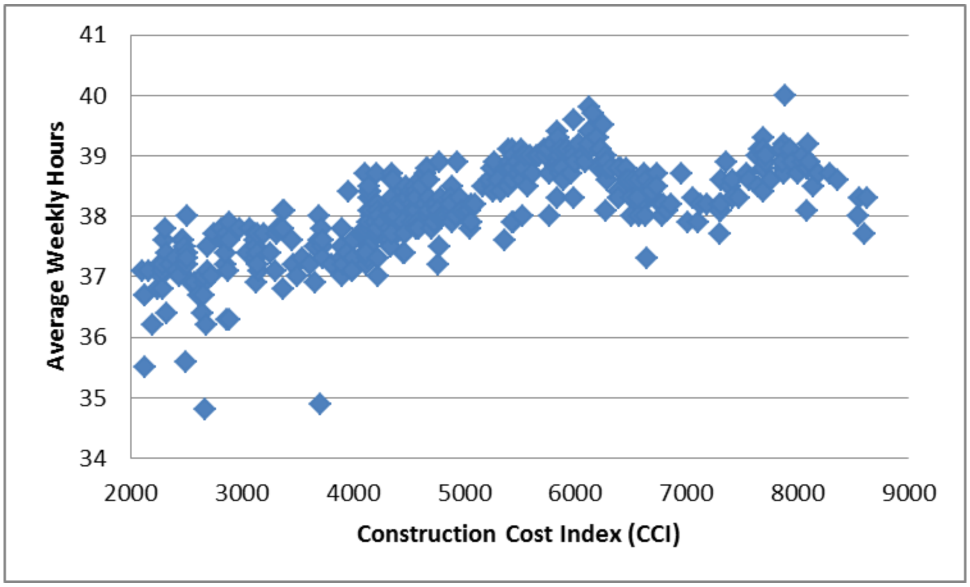


Figure 3.7: Scatter plot representing the relationship between CCI and average weekly hours

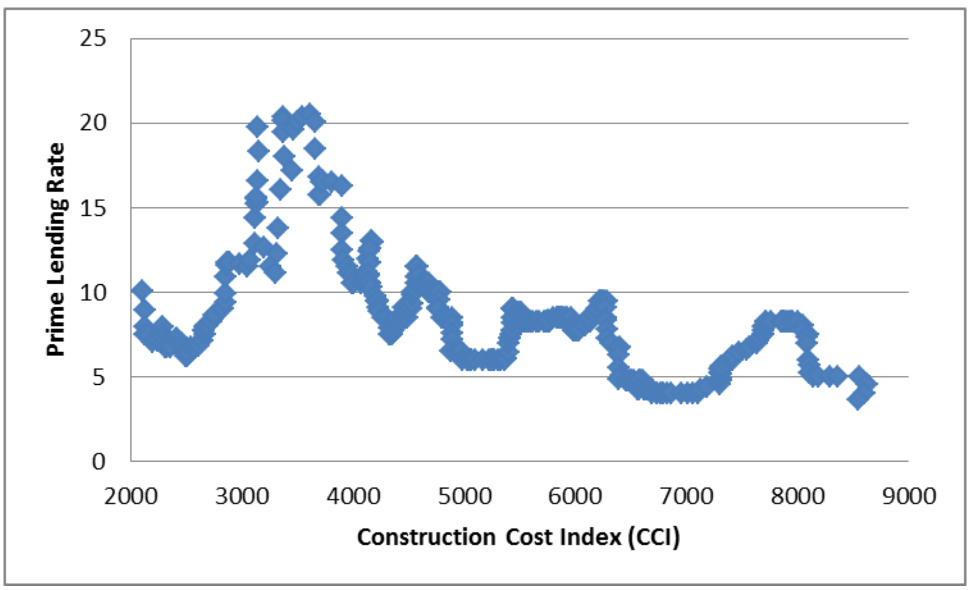


Figure 3.8: Scatter plot representing the relationship between CCI and prime lending rate

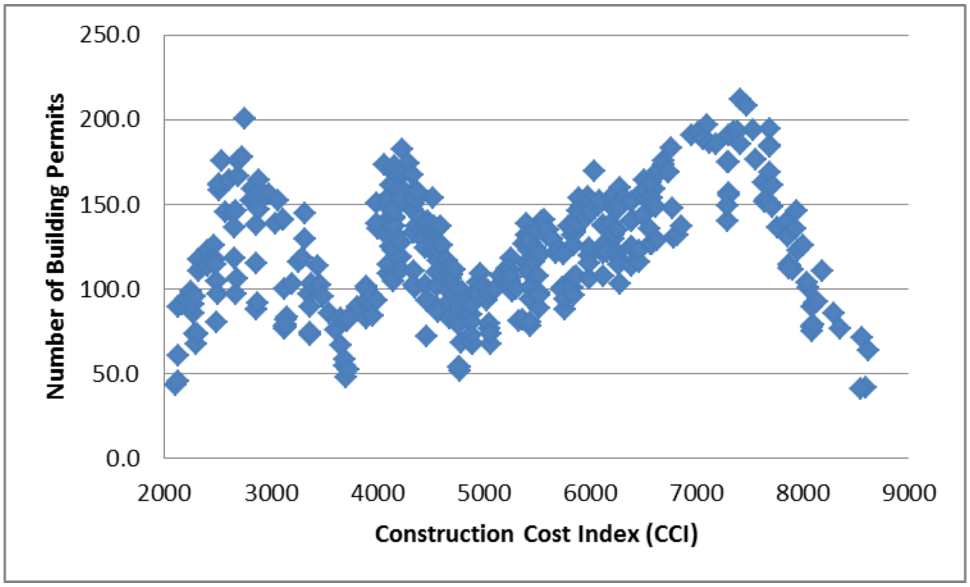


Figure 3.9: Scatter plot representing the relationship between CCI and number of building permits [Thousands of units]

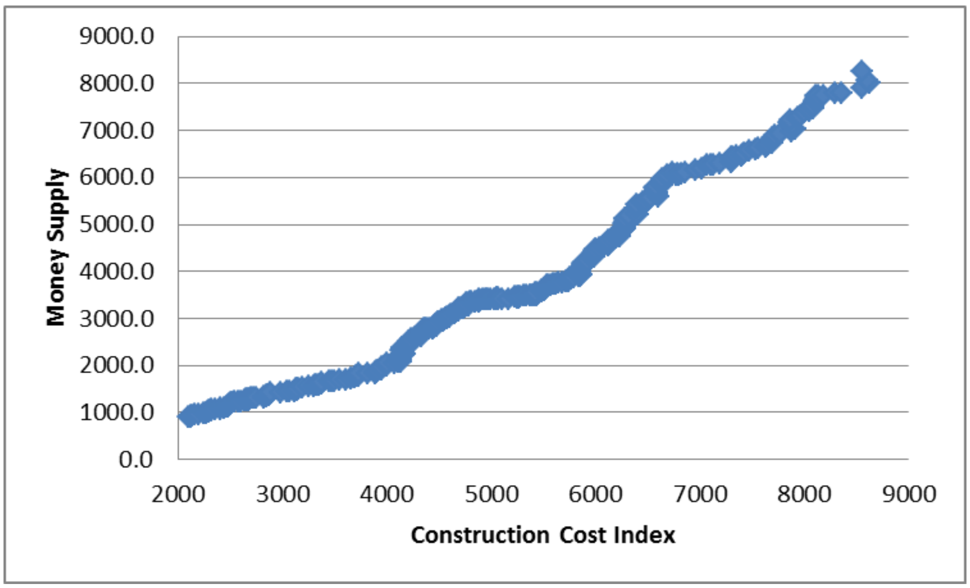


Figure 3.10: Scatter plot representing the relationship between CCI and money supply

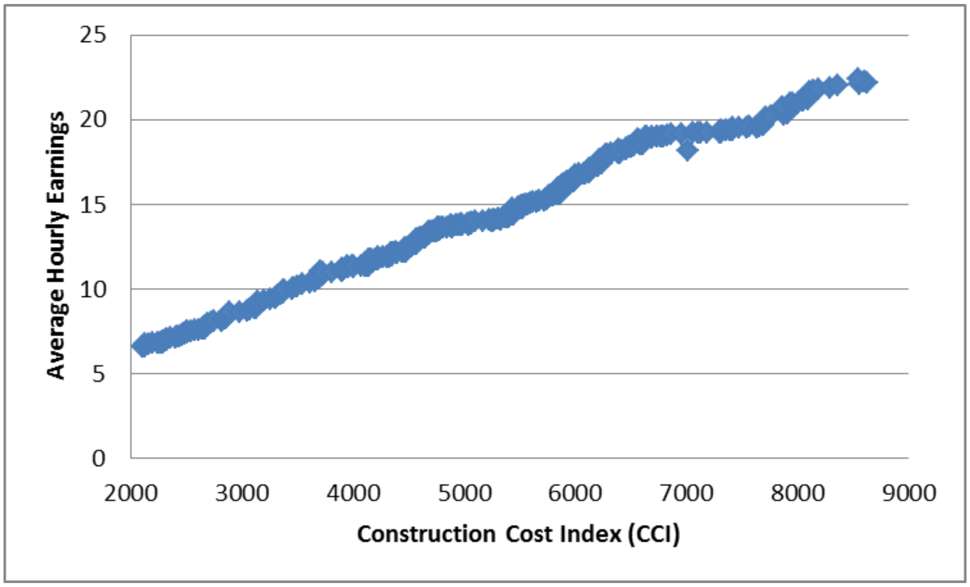


Figure 3.11: Scatter plot representing the relationship between CCI and average hourly earnings

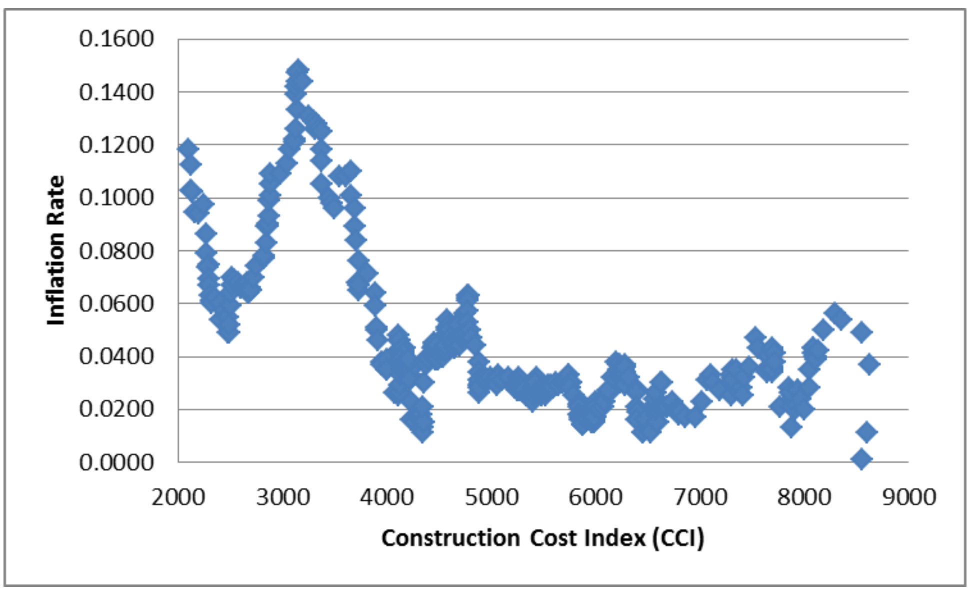


Figure 3.12: Scatter plot representing the relationship between CCI and inflation rate

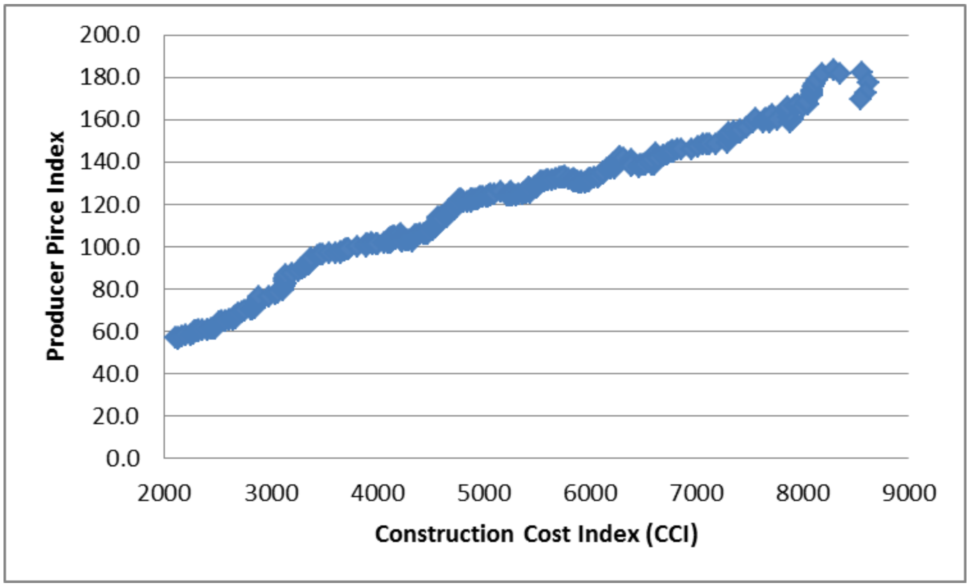


Figure 3.13: Scatter plot representing the relationship between CCI and producer price index

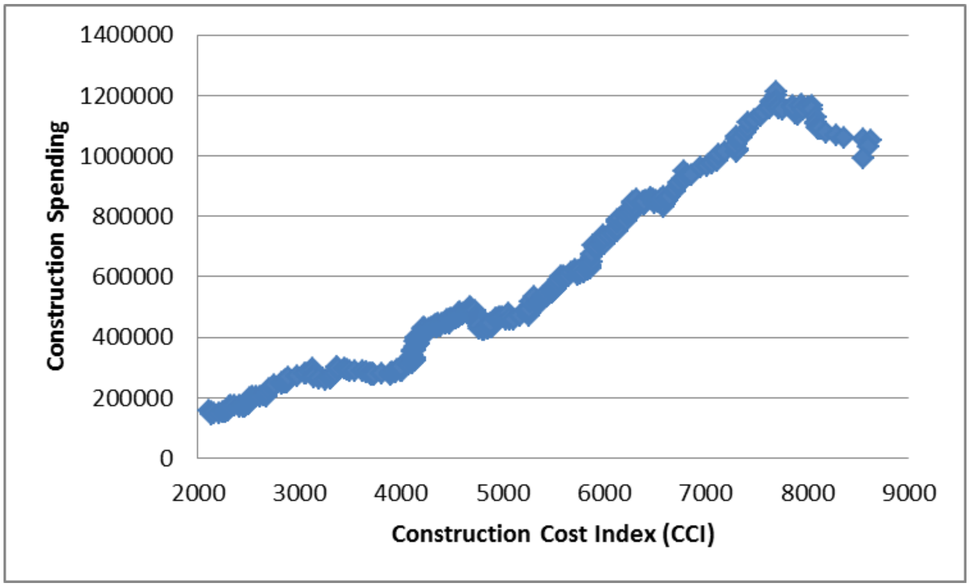


Figure 3.14: Scatter plot representing the relationship between CCI and construction spending [Thousands of units]

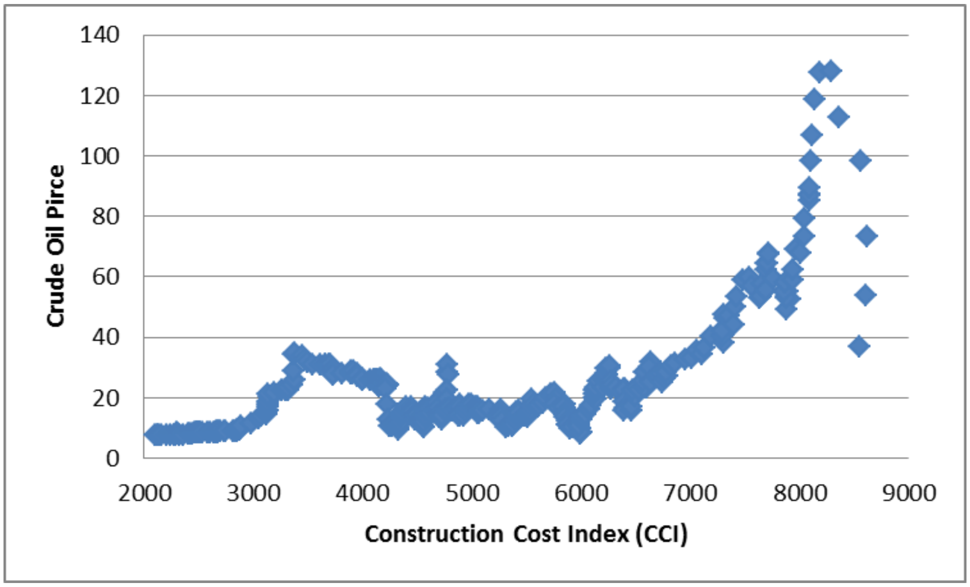


Figure 3.15: Scatter plot representing the relationship between CCI and crude oil price

Unit Root Test for Stationarity

Before Granger causality test is implemented, unit root test is used to identify the order of integration of the variables. The minimum number of times that a time series needs to be differenced for being transformed to stationary time series is the time series' order of integration. Identifying the order of integration precedes statistical tests because variables with the same order of integration can be further used in the Granger causality and cointegration tests. Augmented Dickey-Fuller (ADF) test, proposed by Dickey and Fuller (1979) and extended by Said and Dickey (1984), is used to examine whether the variables are stationary.

Augmented Dickey-Fuller (ADF) test is used to identify the order of integration of the potential leading indicators with a general regression equation:

$$\Delta y_t = \alpha + \delta t + \gamma y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + u_t \qquad (3.1)$$

where Δyt represents the lagged first differences and α , γ , and β are the model parameters that need to be estimated. If y_t is the value of time series y at time t, (y_t-y_{t-1}) represents the lagged first difference (Δy_t) of time series y at time t. p represents the lag length of the test. This test is sensitive to the lag length and selection of the lag length is one of the practical issues related to the ADF test. Akaike Information Criterion (AIC) is used to identify the lag lengths (Akaike, 1974). The null hypothesis is $H_0 : \gamma = 0$; while $H_0 : \gamma < 0$ is the alternative hypothesis. In other words, the null hypothesis is that the time series is not stationary and the alternative hypothesis is that the time series is stationary. Critical values recommended by Banerjee et al. (1993) are used for the unit root test. The more negative the ADF t-statistic is, the stronger the rejection of the null hypothesis (not stationary) at some level of confidence is.

Results of Unit Root Test for Stationarity

The results of ADF unit root tests are presented in Table 3.2. It is shown that all the variables are not stationary. Consumer price index, federal funds rate, unemployment rate, average weekly hours, Prime lending rate, building permits, Dow Jones Industrial Average, crude oil price, producer price index, housing starts, construction spending and CCI become stationary by applying the differencing operator once. The null hypothesis of non-stationarity is rejected at 1% significance level for the differenced consumer price index, federal funds rate, unemployment rate, average weekly hours, Prime lending rate, building permits, Dow Jones Industrial Average, crude oil price, producer price index, housing starts, construction spending and CCI. Therefore, based on definition, these variables are integrated of order 1. Based on these results, Granger causality test is applied to examine whether the first differenced time series of these variables Granger cause the first differenced time series of CCI.

Table 3.2: Results of ADF unit root tests for CCI and the explanatory variables

Variable	**ADF t-statistic**	**Variable**	**ADF t-statistic**
CCI	0.70 (6)	ΔCCI	-8.70 (5) **
CPI	-1.78 (9)	ΔCPI	-4.36 (10) **
FFR	-2.99 (10)	ΔFFR	-6.45 (10) **
UR	-2.85 (7)	ΔUR	-4.52 (5) **
ELC	-2.62 (7)	ΔELC	-3.01 (6)
AWH	-3.27 (6)	ΔAWH	-12.17 (5) **
PLR	-3.07 (9)	ΔPLR	-5.63 (10) **
BP	-0.58 (10)	ΔBP	-11.65 (10) **
MS	2.15 (10)	ΔMS	-1.54 (10)
AHE	-0.92 (10)	ΔAHE	-2.84 (10)
DJIA	-1.85 (1)	ΔDJIA	-14.84 (1) **
COP	-2.08 (6)	ΔCOP	-5.36 (10) **
PPI	-2.59 (10)	ΔPPI	-4.04 (10) **
HS	-0.69 (10)	ΔHS	-11.54 (10) **
CS	-1.90 (7)	ΔCS	-4.40 (6) **
GDP	-3.13 (10)	ΔGDP	-1.66 (10)
GDPIPD	-3.68 (10) *	ΔGDPIPD	-2.66 (10)

*Notes: * indicates rejection of the null hypothesis at the 5% significance level; ** indicates rejection of the null hypothesis at the 1% significance level*

Granger Causality Test for Identifying Leading Indicators of CCI

Granger causality test, which is a statistical hypothesis test, determines whether time series of a variable is useful to predict time series of another variable (Granger 1969). The null hypothesis of the Granger causality test is that the past p values of X do not help to predict Y. p is called lag length of the Granger causality test. The rejection of null hypothesis means that there is enough evidence to state that the past p values of X can be helpful to predict the values of Y. The results of Granger Causality test are sensitive to the chosen lag length (p). Since the Granger causality test is sensitive to the number of lag lengths, it is applied for 6, 12, 18, 24, 30, and 36 lag lengths to examine whether the potential explanatory variables Granger cause CCI. These lag lengths represent a 3-year time horizon which is typically used for examining the predictability of forecasting models of the construction cost (Wong and Ng 2010). A variable is explanatory if the null hypothesis is rejected in at least one of the specified lag lengths. The variable is a consistent explanatory variable if the null hypothesis is rejected at all the specified lag lengths.

The interpretations of the lag length in Granger causality test and the lead-lag relationship can be very important and useful for forecasting applications. Therefore, the rest of this paragraph is dedicated to an example to clarify these concepts in a single example. Suppose the goal is to identify the lead-lag relationship among X and Y. The identification of the lead-lag relationship can start with testing if the past values of X help to predict Y. Suppose the null hypothesis of the Granger causality test is rejected at a given lag length p. This means that the past p values of X are helpful to predict the values of Y. In other words, X leads Y for p values. It is possible to increase p and repeat the test. This test can be repeated until the null hypothesis of Granger causality is not rejected anymore (the lead-lag relationship fades). The same approach can be used to investigate if the past values of Y help to predict X (Y leads X).

Researchers use Granger causality test to study the lead-lag relationship between economic and construction variables. For instance, Granger causality test was used to examine the effects of fluctuations in the money supply on the fluctuations in construction activity flows in Hong Kong (TSE and Raftery, 2001). It was used to evaluate the causal relationship between construction and other economic sectors in Singapore (Lean, 2001). It was applied to study the effects of shocks in construction outputs on major economic indicators in Singapore (Chan, 2002). Wong et al. (2008) used Granger causality test to show how construction outputs (measured by gross value of construction works) drive the economic growth (measured by Gross Domestic Product(GDP)) in Hong Kong. Wong and Ng (2010) used Granger causality test to identify the leading indicators of tender price index in Hong Kong. Granger causality test has not been applied for identifying the leading indicators of CCI.

Bivariate regression models are used to test whether potential variables for leading indicators Granger cause CCI. For instance, the following bivariate regression model is used to test whether crude oil price Granger causes CCI.

$$\Delta CCI_t = \sum_{i=1}^{p} \alpha_i \, \Delta CCI_{t-i} + \sum_{i=1}^{p} \beta_i \, \Delta COP_{t-i} + u_t \tag{3.2}$$

where COP represents crude oil price. COP is Granger causing CCI if $\forall i \ \beta_i = 0$ in Equation 3.2.

Results of Granger Causality Test for Identifying Leading Indicators of CCI

Table 3.3 summarizes the results of Granger causality tests between CCI and the variables in various lag lengths. The results show that consumer price index, crude oil price, and producer price index consistently Granger cause CCI at all the specified lag lengths. Therefore, consumer price index, crude oil price, and producer price index are consistent explanatory variables (leading indicators) of CCI. Building permits and housing starts Granger cause CCI in lower lag lengths. Building permits and housing starts are among the explanatory variables (leading indicators) of CCI.

Summary

Based on the unit root tests and Granger causality tests, consumer price index, crude oil price, producer price index, housing starts and building permits are selected as leading indicators of CCI. In other words, past values of these variables contain information that is useful for forecasting CCI. The trends of these variables are presented in Figure 3.1.

Table 3.3: Results of Granger causality test between CCI and the explanatory variables

	F Statistics					
Null hypothesis	**Lag 6**	**Lag 12**	**Lag 18**	**Lag 24**	**Lag 30**	**Lag 36**
ΔCPI does not Granger cause ΔCCI	7.51**	4.61**	3.61**	3.06**	2.82**	2.81**
ΔFFR does not Granger cause ΔCCI	1.15	1.51	1.25	1.23	1.16	1.10
ΔUR does not Granger cause ΔCCI	1.85	1.06	0.96	0.94	0.83	0.78
ΔAWH does not Granger cause ΔCCI	0.88	0.84	1.05	1.06	1.05	1.01
ΔPLR does not Granger cause ΔCCI	1.29	1.41	1.24	1.22	1.18	1.12
ΔBP does not Granger cause ΔCCI	3.81**	1.99*	1.41	1.15	1.34	1.29
ΔDJIA does not Granger cause ΔCCI	1.35	1.23	1.48	1.44	1.39	1.41
ΔCOP does not Granger cause ΔCCI	6.42**	3.78**	2.86**	2.40**	2.21**	2.19**
ΔPPI does not Granger cause ΔCCI	3.81**	2.11*	1.79*	2.13**	1.80**	1.73**
ΔHS does not Granger cause ΔCCI	4.43**	2.27**	1.46	1.27	1.23	1.26
ΔCS does not Granger cause ΔCCI	0.60	0.54	0.84	1.35	1.24	1.28

*Notes: * indicates rejection of the null hypothesis at the 5% significance level; ** indicates rejection of the null hypothesis at the 1% significance level*

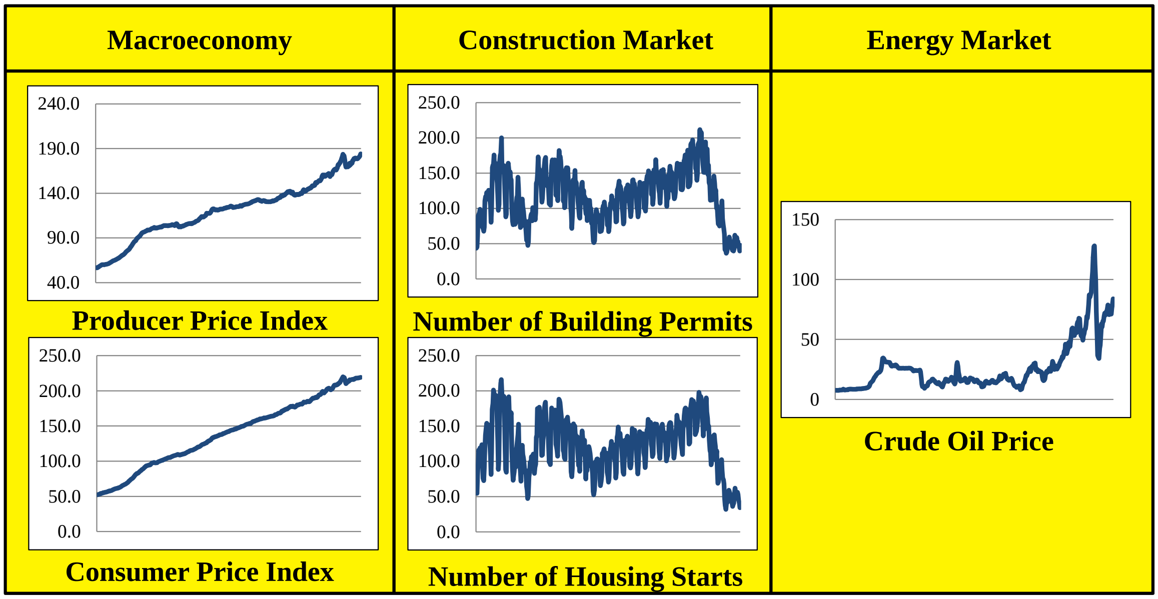


Figure 3.16: Trends of identified leading indicators of CCI

CHAPTER 4

CREATING MULTIVARIATE TIME SERIES MODELS

Chapter 4 shows how appropriate type of multivariate time series models is selected based on the statistical tests. Several multivariate time series models are also created in Chapter 4. These models are validated through comparison with existing univariate time series models for forecasting CCI.

Cointegration test

In order to choose right multivariate time series models, we need to implement a test, known as cointegration test, to examine if CCI and the leading indicators are cointegrated (Pfaff 2008). A group of variables with a specific order of integration are cointegrated if a linear combination of them has a lower order of integration. If a group of variables are cointegrated, they do not drift apart increasingly as time goes and it is possible to conclude that they are related in the long run. In this chapter, a cointegration test proposed by Johansen (1988) and extended by Johansen and Juselius (1990) is used to examine whether CCI is cointegrated with its leading indicators. This test identifies the number of cointegrating relationships denoted by r in a vector of variables. The null hypothesis of this test is that r is less than or equal to a specific value. For instance, if $r \leq 0$ is rejected for a vector of variables, we can conclude that there is at least one cointegrating relationship in the vector. The critical values proposed by Osterwald-lenum (1992) are used for rejection of the null hypothesis.

In order to have a good understanding about the characteristics of the leading indicators, the cointegration test is implemented for three types of vectors. First, it is applied for the vector of CCI and all the leading indicators (i.e., consumer price index, building permits, crude oil price, producer price index, and housing starts). Second, it is applied for the vector of CCI and all the consistent leading indicators (i.e., consumer price index, crude oil price and producer price index). Last, it is applied for bivariate vectors of CCI and each consistent leading indicator.

Results of Cointegration test

The results of Johansen's cointegration test for the vector of CCI and all the leading indicators (i.e., consumer price index, building permits, crude oil price, producer price index, and housing starts) are presented in Table 4.1. Based on the AIC criterion, the lag length of 10 is selected for the test. r represents the number of cointegrating relationships between CCI and the leading indicators. The trace statistics show that null hypothesis of $r \leq 2$ can be rejected at 5% significant level. It is concluded that CCI and all the leading indicators are cointegrated with $r=3$ at the 5% significance level.

Table 4.1: Results of the Johansen cointegration tests for the vector of CCI and all the leading indicators

Null hypothesis	**Trace statistics**	**5% critical value**	**1% critical value**
r = 0	215.11**	102.14	111.01
r ≤ 1	97.97**	76.07	84.45
r ≤ 2	57.43*	53.12	60.16
r ≤ 3	32.42	34.91	41.07
r ≤ 4	16.26	19.96	24.60
r ≤ 5	5.59	9.24	12.97

*Notes: r represents the number of cointegrating relationships; * indicates rejection of the null hypothesis at the 5% significance level; ** indicates rejection of the null hypothesis at the 1% significance level*

Johansen's cointegration test is also applied for the vector of CCI and all the consistent leading indicators (i.e., consumer price index, crude oil price and producer price index). The results of the Johansen's cointegration test indicate that r≤1 can be rejected at 1% significant level. Therefore, CCI and all the leading indicators are cointegrated with r=2 at the 1% significance level. Johansen's cointegration test is also applied for bivariate vectors of CCI and each consistent leading indicator. The results of the Johansen's cointegration test indicate that null hypothesis of no cointegrating relationship (r=0) can be rejected at 1% significant level for the bivariate vectors of CCI and consumer price index, CCI and crude oil price, and CCI and producer price index.

Vector Error Correction Models

Vector Error Correction (VEC) models are recommended for multivariate time series modelling where the variables are cointegrated (Pfaff 2008). Hence, VEC models are created in this study to forecast CCI. The long-run form of VEC model can be represented by the following equation:

$$\Delta y_t = \sum_{i=1}^{p-1} A_i \Delta y_{t-i} + B y_{t-p} + C + \varepsilon_t \qquad (4.1)$$

where y_t is the (K×1) vector of time series at period t, K is the number of variables (e.g., it is 6 if CCI and all the leading indicators are included in the model), A_i (i=1,…,p-1) are (K×K) coefficient matrices of endogenous variables containing the cumulative long-run impacts, B is (K×K) coefficient matrix, C is (K×1) vector of constants, and ε_t is (K×1) vector of error terms.

Equation (4.1) has also the following equivalent representation:

$$y_t = \Pi_1 y_{t-1} + ... + \Pi_p y_{t-p} + C + u_t \qquad (4.2)$$

$$A_i = -(I - \Pi_1 - ... - \Pi_i), \quad for \quad i = 1,.., p-1 \qquad (4.3)$$

$$B = -(I - \Pi_1 - \ldots - \Pi_p) \qquad (4.4)$$

where I is the $(K \times K)$ identity matrix and u_t is $(K \times 1)$ vector of error terms.

Based on the results of statistical tests, five VEC models are created. The first VEC model includes CCI and all the leading indicators. The second VEC model includes CCI and all the consistent leading indicators. The third to fifth VEC models are bivariate VEC models including CCI and each consistent leading indicator. Table 4.2 summarizes the variables in the VEC models and the number of cointegrating relationships.

Table 4.2: VEC models, their variables, and their number of cointegrating relationships

VEC model	**Variables**	**r**
$VECM_1$	CCI,CPI,HS,BP,COP,PPI	3
$VECM_2$	CCI,CPI,PPI,COP	2
$VECM_3$	CCI,PPI	1
$VECM_4$	CCI,CPI	1
$VECM_5$	CCI,COP	1

Notes: VECM is Vector Error Correction Model; r denotes the number of cointegrating relationships; $VECM_1$ represents the model including CCI and all the explanatory variables; $VECM_2$ represents the model including CCI and all the consistent explanatory variables; $VECM_3$ represents the model including CCI and producer price index; $VECM_4$ represents the model including CCI and consumer price index; $VECM_5$ represents the model including CCI and crude oil price

The coefficients of these VEC models are estimated using Gaussian maximum likelihood procedure (Johansen 1995). Equation (4.5) shows the estimated model including CCI and crude oil price.

$$\begin{bmatrix} \text{CCI} \\ \text{COP} \end{bmatrix}_t = \begin{bmatrix} 1.2301 & -0.7266 \\ -0.0118 & 1.6681 \end{bmatrix} \begin{bmatrix} \text{CCI} \\ \text{COP} \end{bmatrix}_{t-1} + \begin{bmatrix} -0.1979 & 2.7593 \\ 0.0165 & -0.6801 \end{bmatrix} \begin{bmatrix} \text{CCI} \\ \text{COP} \end{bmatrix}_{t-2} +$$
$$\begin{bmatrix} -0.0943 & -1.0293 \\ -0.0267 & -0.0384 \end{bmatrix} \begin{bmatrix} \text{CCI} \\ \text{COP} \end{bmatrix}_{t-3} + \begin{bmatrix} 0.0104 & -1.0240 \\ 0.0348 & 0.0277 \end{bmatrix} \begin{bmatrix} \text{CCI} \\ \text{COP} \end{bmatrix}_{t-4} +$$
$$\begin{bmatrix} -0.0678 & 1.5072 \\ -0.009 & -0.1340 \end{bmatrix} \begin{bmatrix} \text{CCI} \\ \text{COP} \end{bmatrix}_{t-5} + \begin{bmatrix} 0.1189 & -1.2923 \\ -0.0034 & 0.1599 \end{bmatrix} \begin{bmatrix} \text{CCI} \\ \text{COP} \end{bmatrix}_{t-6} + \begin{bmatrix} 13.0903 \\ 0.2164 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}_t \tag{4.5}$$

Diagnostic tests are used to examine lack of serial correlation and constant variance (lack of heteroskedasticity) among the residuals. Existence of serial correlation among the residuals of the VEC models is investigated using Breusch–Godfrey serial correlation Lagrange multiplier (LM) test proposed by Breusch (1978) and Godfrey (1978). The null hypothesis of Breusch-Godfrey test is that there is no serial correlation in the residuals. ARCH test (Engle 1982) is applied for investigating heteroskedasticity in the residuals of CCI. Constant variance of the residuals is the null hypothesis of ARCH test. Table 4.3 summarizes the results of Breusch-Godfrey test on the residuals of VEC models. It also shows the results of ARCH test on the CCI residuals of the VEC models.

Table 4.3: Results of Breusch-Godfrey LM and ARCH tests for the residuals of VEC models

Model	Breusch-Godfrey test statistics	ARCH statistics
$VECM_1$	688.8**	11.4
$VECM_2$	330.5**	11.2
$VECM_3$	36.0	18.5**
$VECM_4$	42.52*	16.2*
$VECM_5$	30.3	11.8

*Notes: * indicates rejection of the null hypothesis at the 5% significance level; ** indicates rejection of the null hypothesis at the 1% significance level; VECM is Vector Error Correction Model; $VECM_1$ represents the model including CCI and all the explanatory variables; $VECM_2$ represents the model including CCI and all the consistent explanatory variables; $VECM_3$ represents the model including CCI and producer price index; $VECM_4$ represents the model including CCI and consumer price index; $VECM_5$ represents the model including CCI and crude oil price*

The results of Breusch-Godfrey test indicates that the residuals of the VEC model including CCI and all the leading indicators ($VECM_1$) and the VEC model including CCI and all the consistent leading indicators ($VECM_2$) have serial correlation. The results show that the null hypothesis (no serial correlation) cannot be rejected for the VEC model including CCI and producer price index ($VECM_3$) and the VEC model including CCI and crude oil price ($VECM_5$) at *1%* significance level. The null hypothesis (no serial correlation) can be rejected at *5%* significance level for the VEC model including CCI and consumer price index ($VECM_4$).

The results of ARCH test show that the residuals of CCI have constant variance in the VEC model including CCI and all the leading indicators ($VECM_1$), the VEC model including CCI and all the consistent leading indicators ($VECM_2$), and the VEC model including CCI and crude oil price ($VECM_5$). The null hypothesis of constant variance of the CCI residuals can be rejected at *5%* and *1%* significance levels for the VEC model including CCI and consumer price index ($VECM_4$) and the VEC model including CCI and producer index ($VECM_3$), respectively.

Predictability of the VEC Models

The predictability of the VEC models is compared with the predictability of the existing univariate time series models for forecasting CCI. Ashuri and Lu (2010) 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 models in terms of the predictability. The predictability of the time series models are compared based on two error measures: Mean Absolute Prediction Error (MAPE) and Mean Squared Error (MSE). The following equations show how these error measures are calculated.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\hat{Y}_t - Y_t\right|}{Y_t}\times 100 \tag{3.6}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_t - Y_t)^2 \tag{3.7}$$

where $\hat{Y}_t$ is the forecasted CCI by time series models, Y_t is the actual CCI and n is the total number of forecasted data points. Table 4.4 presents MAPE and MSE calculated using forecasted data points from the models and the testing data from January 2009 to December 2011. Since training data from January 1975 to December 2008 are used for the estimation of the models, the predictions represent three-year-ahead CCI. Based on the results shown in Table 4.4, bivariate VEC models ($VECM_3$, $VECM_4$, and $VECM_5$) provide better forecasts (less MAPE and MSE) than the VEC model including CCI and all the leading indicators ($VECM_1$) and the VEC model including CCI and all the consistent leading indicators ($VECM_2$). Two Bivariate VEC models (the VEC model including CCI and producer price index and the VEC model including CCI and crude oil price) provide better forecasts than the univariate time series models. Among bivariate VEC models, the model including CCI and producer price index provides the best forecasts.

Table 4.4: Forecasting errors of time series models

Measure	$VECM_1$	$VECM_2$	$VECM_3$	$VECM_4$	$VECM_5$	S-ARIMA	HW-ES
MAPE	4.93%	3.58%	0.84%	1.48%	0.96%	1.40%	2.68%
MSE	218736.9	115313.6	7600.1	20594.0	10544.9	17921.6	86890.7

Notes: VECM represents Vector Error Correction Model; S-ARIMA represents Seasonal ARIMA; HW-ES represents Holt-Winters Exponential Smoothing; AR represents Autoregressive; VECM1 represents the VECM including CCI and all the explanatory variables; VECM2 represents the VECM including CCI and all the consistent explanatory variables; VECM3 represents the VECM including CCI and producer price index; VECM4 represents the VECM including CCI and consumer price index; VECM5 represents the VECM including CCI and crude oil price; MAPE represents Mean Absolute Prediction Error; MSE represents Mean Squared Error

It is an interesting result that the bivariate VEC models provide better forecasts than the VEC models including all the consistent leading indicators. This result shows that adding information from all the consistent leading indicators do not necessarily improve the predictability of multivariate time series models. From the theoretical stand point, there is no theory to express that all the explanatory indicators should be used within multivariate time series models. In fact, adding more data can increase the chance of overfitting. Furthermore, the consistent leading indicators might also be correlated and this correlation might interfere with the forecasting capabilities of the multivariate time series models. These reasons motivated us to test different multivariate time series model with various number of leading indicators in order to find proper time series models. Appendix A provides all the multivariate time series models that can be created with the CCI and its five identified leading indicators. The results show that there are several VEC models that are more accurate than the univariate time series models for forecasting CCI. VEC models with the following combinations of CCI and leading indicators provide better out-of-sample forecasts than existing univariate models:

- CCI, COP
- CCI, PPI
- CCI, HS, COP
- CCI, HS, PPI
- CCI, BP, COP
- CCI, BP, PPI
- CCI, COP, PPI

Although these results show that there are several VEC models that provide better forecasts than existing univariate models, the results of diagnosis tests in the Appendix show that the VEC model including CCI and COP is the only multivariate time series model that pass diagnosis tests.

CHAPTER 5

TESTING PREDICTABILITY OF MULTIVARIATE MODELS FOR FORECASTING CCI USING SIMULATED DATA

Purpose

The purpose of this chapter is to simulate crude oil price and CCI and show that multivariate time series modeling provides better solution than univariate time series modeling for forecasting simulated CCI. Crude oil price is selected for simulation because the research shows its superiority as the leading indicator for forecasting CCI (See Chapter 4).

Approach

The following steps are taken to test the predictability of multivariate time series models for forecasting CCI using simulated data:

- Create 50 paths of crude oil price using Geometric Brownian Motion
- Create 50 paths of CCI using Gaussian Process that is considering the relationship between CCI and crude oil price over time
- Create 50 multivariate and univariate time series models using the simulated data and compare the predictability of univariate and multivariate time series models

Crude Oil Price Simulation

Two stochastic processes have often been used in the literature to simulate crude oil price: Geometric Brownian Motion (GBM) and Mean-Reverting Process. Based on recent studies (Chikobvu 2010; Gemen 2007), GBM is used in this research to simulate crude oil price. The data up to 2000 is used for estimating parameters of GBM. The data from 2001 to 2011 is generated using GBM. Figure 5.1 shows the 50 path created using GBM.

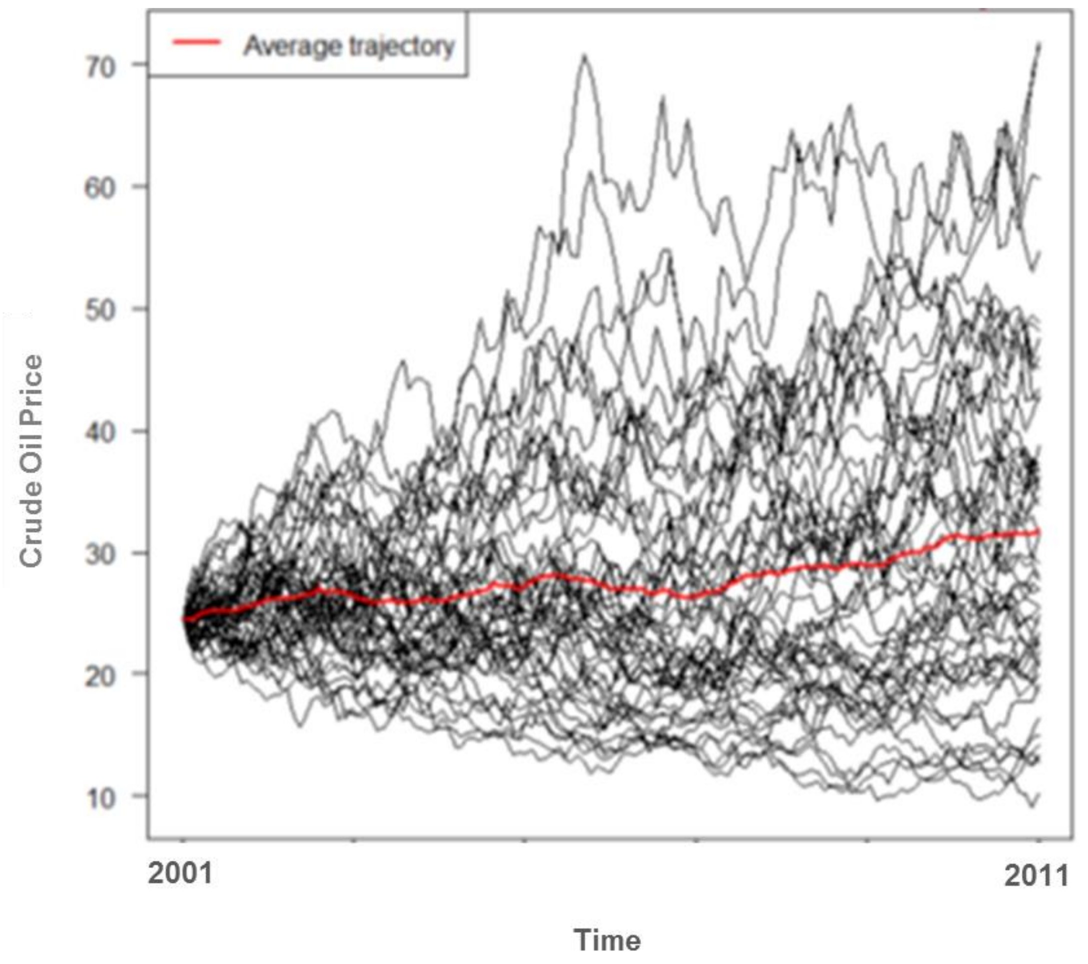


Figure 5.1: 50 paths of crude oil price simulated using GBM

CCI Simulation

50 paths of CCI are created using Gaussian Process (Dancik 2007; Dancik and Dorman 2008) that is considering the relationship between CCI and COP over time. For creating each path, suppose $[CCI(COP^{(1)}),...,CCI(COP^{(m)})]$ is the vector of observed CCI for each Crude oil price. The correlation between any two observed CCI is assumed to have the product exponential form that is a popular choice (Qian et al. 2008).

$$C(\beta)_{i,j} \equiv cor\left(CCI(COP^{(i)}),CCI(COP^{(j)})\right) = \exp\left(-\beta(COP^{(i)} - COP^{(j)})^2\right) \tag{5.1}$$

The vector of CCI is distributed according to the following multivariate normal distribution:

$$CCI \sim MVN(M, \sigma_{GP}^2 C(\beta) + \sigma_e^2 I) \tag{5.2}$$

Where M is the unconditional mean of observed CCI and σ_{GP}^2 is the unconditional variance of an expected CCI and σ_e^2 is variance due to the stochasticity of the CCI (e.g., random noise). The parameters of the Gaussian process are estimated using Maximum Likelihood Estimation (Dancik 2007).

Comparing Predictability of Univariate and Multivariate Time Series Models Using Simulated Data

50 paths of crude oil price and 50 corresponding paths of CCI are created so far. These paths are used to create multivariate and univariate time series models. The data up to December 2008 is used for training multivariate time series model (VEC model including CCI and crude oil price) and univariate time series model (Seasonal ARIMA). Note that this data contains simulated data. The data from January 2009 to December 2011 is used for testing predictability. The average of MAPE and MSE for the multivariate and univariate time series models is presented in Table 5.1. This figure shows that the multivariate modeling is more accurate than univariate modeling for forecasting CCI.

Table 5.1: Forecasting errors of time series models for predicting simulated CCI

	VECM including CCI and COP	Seasonal ARIMA
MAPE	1.27%	1.70%
MSE	13313.7	18495.8

Sensitivity Analysis

Here, different amounts of errors are added to the simulated data in order to investigate how sensitive the models are. Errors are modeled by sampling from a normal distribution with mean zero and standard deviation equal to x percent of the simulated value and adding that to the simulate data. The following sub sections show the predictability of VEC including CCI and COP and Seasonal ARIMA where different amounts of errors are added to the simulated data.

Adding 1% Error

Table 5.2 shows the predictability of VEC including CCI and COP and Seasonal ARIMA where 1% error is added to the simulated data. The results show that the accuracy is slightly less than the models without any error. The results show that the VEC model including CCI and COP is more accurate than Seasonal ARIMA.

Table 5.2: Forecasting errors of time series models for predicting simulated CCI

	VECM including CCI and COP	Seasonal ARIMA
MAPE	1.59%	2.31%
MSE	17160.6	27858.7

Adding 5% Error

Table 5.3 shows the predictability of VEC including CCI and COP and Seasonal ARIMA where 5% error is added to the simulated data. The results show that the accuracy is less than the models without any error. The results show that the VEC model including CCI and COP is more accurate than Seasonal ARIMA.

Table 5.3: Forecasting errors of time series models for predicting simulated CCI

	VECM including CCI and COP	Seasonal ARIMA
MAPE	4.58%	5.91%
MSE	104396.5	164845.8

Adding 10% Error

Table 5.4 shows the predictability of VEC including CCI and COP and Seasonal ARIMA where 10% error is added to the simulated data. The results show that the accuracy is highly less than the models without any error. The results show that the VEC model including CCI and COP is more accurate than Seasonal ARIMA.

Table 5.4: Forecasting errors of time series models for predicting simulated CCI

	VECM including CCI and COP	Seasonal ARIMA
MAPE	8.97%	10.77%
MSE	369731.8	526426.9

CHAPTER 6

FORECASTING NATIONAL HIGHWAY CONSTRUCTION COST INDEX

Federal Highway Administration (FHWA) has published National Highway Construction Cost Index (NHCCI) quarterly since 2003. NHCCI measures the average change over time in the roadway construction material and service prices paid by State transportation departments (FHWA 2014). This index is used to track changes in highway construction costs, and to help cost estimators to take into account highway construction cost changes over time. Accurate forecasting of NHCCI is critical for preparing accurate bids and prevents under- or over-estimation of costs of highway projects. However, NHCCI is subject to significant variations that make forecasting challenging.

Accurate forecasting of NHCCI requires knowledge about historical variations of highway construction costs and temporal relationships of these variations with fluctuations in some explanatory variables representing macroeconomic and construction market conditions. However, existing literature lacks rigorous identification of explanatory variables that are useful to predict NHCCI. Explanatory variables that are useful to forecast NHCCI are the leading indicators of NHCCI. These explanatory variables are called leading indicators because they contain information that is useful to predict future values of NHCCI. These leading indicators should be included in appropriate multivariate models for forecasting NHCCI. These multivariate models should take into account the temporal relationship between NHCCI and the proven leading indicators in order to forecast NHCCI accurately. Existing models do not take into account both proven leading indicators of NHCCI and the temporal relationship among variables.

The objective of this chapter is to identify the leading indicators of NHCCI, and create multivariate statistical models for improving the accuracy of forecasting NHCCI through utilizing information available from the identified leading indicators. The following steps are taken to identify the leading indicators of NHCCI, and create appropriate multivariate statistical models.

- Use statistical tests to identify leading indicators of NHCCI from the pool of the potential leading indicators.
- Use cointegration tests to investigate appropriate type of multivariate time series models to forecast NHCCI.
- Create appropriate multivariate time series models to forecast NHCCI.
- Diagnose the created multivariate time series models.
- Validate results through testing the out-of-sample predictability of the created multivariate time series models.

Statistical Tests to Identify Leading Indicators of NHCCI

Unit Root Test

It is critical to identify whether the variables are stationary before using any other multivariate time series tests because most of these multivariate time series tests, such as Granger Causality tests can only be applied on stationary time series. Unit root tests, such as Augmented Dickey-Fuller (ADF) test are used to test whether variables are stationary. ADF test proposed by Dickey and Fuller (1979) and extended by Said and Dickey (1984) is used in this study to identify whether the variables are stationary. The null hypothesis of this test is that the time series is not stationary and the alternative hypothesis is that the time series is stationary. Critical values recommended by Banerjee et al. (1993) are used for this test.

Results of Unit Root Test

Table 6.1 shows the results of ADF test for NHCCI and the macroeconomic, construction, and energy market variables. It shows that NHCCI is not stationary (The null hypothesis cannot be rejected). Most of the variables (except federal funds rate, employment level in construction, average weekly hours, prime loan rate, and producer price index) are not stationary. Since NHCCI and most of the variables are not stationary, they are differenced and the unit root tests are repeated for the differenced terms. NHCCI, consumer price index, average weekly hours, average hourly earnings, Dow Jones industrial average, and crude oil price become stationary by differencing. The number of times that a variable must be differenced until stationary time series data are created is the order of integration for the variable. The Granger causality test is used for the stationary time series (i.e., differenced time series of NHCCI, consumer price index, average weekly hours, average hourly earnings, Dow Jones industrial average, and crude oil price).

Table 6.1: Results of ADF unit root tests for potential leading indicators of NHCCI

Variable	**Lag Order**	**P-Value**	**Variable**	**Lag Order**	**P-Value**
NHCCI	2	0.63	ΔNHCCI	2	0.07*
CPI	3	0.71	ΔCPI	2	0.01**
FFR	4	0.07*	ΔFFR	3	-1.77
UR	2	0.52	ΔUR	1	-1.78
ELC	4	0.07*	ΔELC	1	-1.37
AWH	10	0.04**	ΔAWH	1	-5.21**
PLR	4	0.07*	ΔPLR	3	-1.75
BP	5	0.69	ΔBP	4	-1.69
MS	1	0.67	ΔMS	1	-3.13
AHE	1	0.46	ΔAHE	1	-3.41*
DJIA	6	0.43	ΔDJIA	1	-3.42*
COP	4	0.42	ΔCOP	4	-3.40*
PPI	1	0.07*	ΔPPI	4	-2.71
HS	5	0.57	ΔHS	4	-1.40
CS	10	0.26	ΔCS	1	-1.97
GDP	2	0.63	ΔGDP	1	-2.53
GDPIPD	2	0.76	ΔGDPIPD	1	-2.90

*Notes: Δ is the first difference operator; * Rejection of the null hypothesis at the 10% significance level; ** Rejection of the null hypothesis at the 5% significance level; Lag orders are selected based on the Akaike Information Criterion (AIC) (Akaike 1974).*

Granger Causality Test

Granger causality test is a multivariate time series test for identifying if one time series data is useful to forecast another one (Granger 1969). If the time series X is useful to forecast time series Y, X Granger causes (leads) Y. Bivariate regression models are used to test whether the variables Granger cause (or lead) NHCCI. For example, the following bivariate regression model is used to test whether consumer price index Granger causes NHCCI.

$$\Delta NHCCI_t = \sum_{i=1}^{p} \alpha_i \Delta NHCCI_{t-i} + \sum_{i=1}^{p} \beta_i \Delta CPI_{t-i} + u_t \qquad (6.1)$$

where CPI represents consumer price index. CPI is Granger causing NHCCI if $\forall i \; \beta_i$. The null hypothesis of the test is that the past p values of CPI do not help the predictability of NHCCI. The rejection of null hypothesis means that the past p values of CPI can be helpful to predict the values of NHCCI.

Results of Granger Causality Test

Table 6.2 shows the results of Granger causality tests. It shows that the null hypotheses (the variables do not Granger cause NHCCI) can be rejected for two variables: average hourly earnings and crude oil price. Therefore, average hourly earnings and crude oil price are the leading indicators of NHCCI and will be used in the following sections to create appropriate multivariate time series models.

Table 6.2: Results of Granger causality tests

Null Hypothesis	Lags				
	2	4	6	8	10
ΔCPI does not Granger cause ΔNHCCI	1.65	2.09	1.30	0.85	0.69
ΔAWH does not Granger cause ΔNHCCI	1.90	1.70	1.34	0.74	3.15
ΔAHE does not Granger cause ΔNHCCI	4.33**	2.32*	1.17	1.88	5.32*
ΔDJIA does not Granger cause ΔNHCCI	0.56	0.53	0.28	0.22	0.29
ΔCOP does not Granger cause ΔNHCCI	8.30**	4.06**	2.43*	1.15	0.83

*Note: Δ is the first difference operator; * Rejection of the null hypothesis at the 10% significance level; ** Rejection of the null hypothesis at the 5% significance level;*

Cointegration Tests to Identify Appropriate Type of Multivariate Time Series Models

Cointegration test can be used to choose the right type of multivariate time series model (Pfaff 2008). A cointegration test proposed by Johansen (1988) and extended by Johansen and Juselius (1990) is implemented to examine whether NHCCI is cointegrated with the identified leading indicators. This cointegration test identifies the number of cointegrating relationships (r) in a group of variables. The null hypothesis of this test is that the number of cointegrating relationships is less than or equal to a specific value (less than the number of variables). The critical values proposed by Osterwald-lenum (1992) are used as thresholds for rejecting the null hypothesis.

The cointegration test is implemented for three types of vectors. First, it is applied for the vector of NHCCI and all the leading indicators (i.e., average hourly earnings and crude oil price). Second, it is applied for the bivariate vector of NHCCI and average hourly earnings. Last, it is applied for the bivariate vector of NHCCI and crude oil price.

Results of Cointegration Tests to Identify Appropriate Type of Multivariate Time Series Models

The results of Johansen's cointegration test for the vector of NHCCI, average hourly earnings, and crude oil price are presented in Table 6.3. The trace statistics show that null hypothesis of $r \leq 2$ can be rejected at 1% significant level. Therefore, it is concluded that NHCCI, average hourly earnings, and crude oil price variables are cointegrated.

Table 6.3: Results of the Johansen cointegration tests for the vector of NHCCI, AHE, and COP

Null hypothesis	**Trace statistics**
r = 0	109.98**
r ≤ 1	53.05**
r ≤ 2	15.34**

*Notes: r represents the number of cointegrating relationships; * indicates rejection of the null hypothesis at the 5% significance level; ** indicates rejection of the null hypothesis at the 1% significance level; AIC criterion is used to choose the lag length for the test.*

The results of Johansen's cointegration test for the vector of NHCCI and average hourly earnings are presented in Table 6.4. The trace statistics show that null hypothesis of r≤1 can be rejected at 1% significant level. Therefore, it is concluded that NHCCI and average hourly earnings are cointegrated.

Table 6.4: Results of the Johansen cointegration tests for the vector of NHCCI and AHE

Null hypothesis	Trace statistics
r = 0	58.64**
r ≤ 1	23.28**

*Notes: r represents the number of cointegrating relationships; * indicates rejection of the null hypothesis at the 5% significance level; ** indicates rejection of the null hypothesis at the 1% significance level; AIC criterion is used to choose the lag length for the test.*

The results of Johansen's cointegration test for the vector of NHCCI and crude oil price are presented in Table 6.5. The trace statistics show that null hypothesis of r≤1 can be rejected at 1% significant level. Therefore, it is concluded that NHCCI and average hourly earnings are cointegrated.

Table 6.5: Results of the Johansen cointegration tests for the vector of NHCCI and COP

Null hypothesis	Trace statistics
r = 0	54.93**
r ≤ 1	16.91**

*Notes: r represents the number of cointegrating relationships; * indicates rejection of the null hypothesis at the 5% significance level; ** indicates rejection of the null hypothesis at the 1% significance level; AIC criterion is used to choose the lag length for the test.*

Vector Error Correction (VEC) models are recommended as the proper time series models for the cointegrated variables (Pfaff 2008). Next section explains how VEC models are created for NHCCI and various combinations of the identified leading indicators.

Vector Error Correction Models

Equation 6.2 shows the VEC model that is used for creating multivariate times series models in this study.

$$\Delta y_t = \sum_{i=1}^{p-1} \Phi_i \Delta y_{t-i} + \Pi y_{t-p} + C + \varepsilon_t \qquad (6.2)$$

y_t is the (N×1) vector of time series at period t (N is the number of variables), Φ_i (i=1,…,p-1) are (N×N) coefficient matrices of endogenous variables, Π is (N×N) coefficient matrix, C is (N×1) vector of constants, and ε_t is (N×1) vector of error terms.

Based on the results of statistical tests, three VEC models are created. The first VEC model ($VECM_1$) includes NHCCI, average hourly earnings and crude oil price. The second VEC model ($VECM_2$) includes NHCCI and average hourly earnings. The third VEC model ($VECM_3$) includes NHCCI and crude oil price. The coefficients of these VEC models are estimated using Gaussian maximum likelihood procedure (Johansen 1995).

Diagnosis Tests

The residuals of the multivariate models created in the previous section should not be serially correlated. Moreover, the variance of the residuals of the multivariate models should be constant. Two diagnostic tests are used to examine whether the residuals of the models follow these modelling assumptions. First, the Breusch–Godfrey serial correlation Lagrange multiplier test (Breusch 1978; Godfrey 1978) is used to test the serial correlation among the residuals of the models. The null hypothesis of Breusch-Godfrey test is that there is no serial correlation in the residuals. Table 6.6 shows the results of Breusch-Godfrey tests on the residuals of the models. According to Table 6.6, the model including NHCCI and average hourly earnings and the model including NHCCI and crude oil price pass the serial correlation test.

Table 6.6: Results of Breusch-Godfrey LM tests for the residuals of VEC models

Model	Breusch-Godfrey test statistics
VECM1	87*
VECM2	52
VECM3	52

*Notes: * indicates rejection of the null hypothesis at the 5% significance level; VECM is Vector Error Correction Model; VECM1 represents the model including NHCCI and all the leading indicators; VECM2 represents the model including NHCCI and average hourly earnings; VECM3 represents the model including NHCCI and crude oil price.*

Second, the Autoregressive Conditional Heteroskedasticity (ARCH) test (Engle 1982) is used to test whether residuals of the multivariate models have constant variance. The null hypothesis of ARCH test is the constant variance of the residuals of the models. Table 6.7 shows the results of the ARCH test on the residuals of the models. According to Table 6.7, the null hypothesis is not rejected for any of the models and all the models pass the constant variance test.

Table 6.7 Results of ARCH tests for the residuals of VEC models

Model	**ARCH statistics**
VECM1	132
VECM2	48
VECM3	48

*Notes: * indicates rejection of the null hypothesis at the 5% significance level; VECM is Vector Error Correction Model; VECM1 represents the model including NHCCI and all the leading indicators; VECM2 represents the model including NHCCI and average hourly earnings; VECM3 represents the model including NHCCI and crude oil price.*

Validation through Testing Out-Of-Sample Predictability

Multivariate time series models are compared with the univariate time series models in order to validate the hypothesis that the leading indicators of NHCCI are useful for accurate forecasting. The 2012 NHCCI data is used for testing out-of-sample predictability of the models. This is called out-of-sample since 2012 data is not used for conducting statistical tests or creating multivariate models in the previous sections (The data from 2003 to 2011 is used for conducting the statistical tests or creating the multivariate models). The out-of-sample predictability of the multivariate time series models is compared with the out-of-sample predictability of two best univariate time series models for forecasting construction cost (Ashuri and Lu 2010): seasonal autoregressive integrated moving average model and Holt-Winters Exponential Smoothing (HW-ES) model. This comparison reveals the value of using historical values in multiple time series with the value of just using NHCCI time series for forecasting NHCCI. The out-of sample predictability of the multivariate time series models and the univariate time series models are reported based on two typical error measures: Mean Absolute Prediction Error (MAPE) and Mean Squared Error (MSE).

Table 6.8 shows the results of testing out-of-sample predictability of the multivariate models versus the univariate models. According to Table 6.8, the multivariate model including NHCCI, crude oil price and average hourly earnings provides the least prediction error (best forecasts). The error of this multivariate model that includes all the identified leading indicators is less than half of the errors of the univariate models. Therefore, the leading indicators of NHCCI are useful for accurate forecasting.

Table 6.8 Out-of-sample predictability of VEC models versus univariate time series models

Measure	VEC model (NHCCI, COP and AHE)	VEC model (NHCCI and AHE)	VEC model (NHCCI, COP)	Seasonal ARIMA	HW-ES
MAPE	2.07%	10.78%	2.73%	4.43%	4.55%
MSE	0.007	0.0191	0.0013	0.0027	0.0033

Summary

Crude oil price and average hourly earnings are the leading indicators of NHCCI selected from the pool of potential leading indicators using statistical tests. Any combination of these leading indicators and NHCCI is cointegrated. Hence, VEC models are the appropriate type of multivariate time series models for forecasting NHCCI using the identified leading indicators. Based on the results of statistical tests, three VEC models are created: $VECM_1$ that includes NHCCI, average hourly earnings and crude oil price, $VECM_2$ that includes NHCCI and average hourly earnings and VECM3 that includes NHCCI and crude oil price. The coefficients of these VEC models are estimated using Gaussian maximum likelihood procedure. $VECM_2$ and $VECM_3$ pass the diagnosis tests and $VECM_1$ and $VECM_3$ provide better out-of-sample forecasting than the univariate time series models.

These findings contribute to the body of knowledge in NHCCI forecasting by rigorous identification of NHCCI leading indicators and creation of multivariate time series models that are more accurate than the univariate time series models for forecasting NHCCI. It is expected that this work contributes to the construction engineering and management community by helping highway cost engineers and investment planners prepare more accurate bids, cost estimates, and budgets for highway projects.

CHAPTER 7

DISCUSSION OF RESULTS

Impact of Crude Oil Price Fluctuations on Construction Cost variations

Crude oil price was found as the only common leading indicator of both CCI and NHCCI. It was shown that crude oil price provides significantly useful information to predict the future values of CCI and NHCCI. Here, the impact of crude oil price fluctuations on construction cost variations is discussed.

Although the existing literature lacks a rigorous empirical analysis indicating the temporal relationship between oil price fluctuations and construction cost variations, several articles and industry reports highlight the impact of fluctuations of oil price on construction costs. For example, Damnjanovic and Zhou (2009) showed that the cost of highway construction is affected by the cost of crude oil. They highlighted this impact on the costs of both directly related construction items, such as asphalt cement and the cost of other construction items, such as concrete cement or construction operations. Wilmot and Cheng (2003) also recognized that the increase in cost of petroleum products as the major cause of the increase in construction costs. Gallagher and Riggs (2006) highlighted existence of a direct relationship between the cost of construction, materials, and oil price.

Olatunji (2010) showed that there is a strong relationship between changes in oil prices and fluctuations in construction costs. 48% of surveyed managers of public educational, transportation and municipal entities identified oil and gas prices as the key factor behind higher construction bids (Reid 2005). Mendell (2006) highlighted the significant impact of energy costs on costs of petroleum-based materials, such as PVC water and sewer pipes and costs of transport costs. Sicotte and Glitman (2011) showed that crude oil prices significantly impact construction costs using regression analysis.

This dissertation provides rigorous empirical evidence showing the significant long-term impact of crude oil price on construction costs.

Explanations Supporting the Observed Impact of Crude Oil Price Fluctuations on Construction Cost Variations

Here, some explanations supporting the observed impact of fluctuations of oil price on construction costs are provided:

- Contractors and suppliers are more cautious when bidding in a volatile oil market. Hence, they include higher contingencies in the projects. Therefore, the volatility of oil price propagates to the construction industry (Damnjanovic and Zhou, 2009).
- The cost of crude oil price affects directly the cost of some construction materials, such as asphalt cements since it is one of their main components (Akimovs, 2013; Damnjanovic and Zhou, 2009; Harmon 2003; Mendell 2006).
- The effects of volatile oil prices go beyond the price of asphalt materials by affecting the costs of gasoline and diesel fuel and consequently impacting the costs of construction operations that are all energy consuming (Davenport 2008; Akimovs, V. 2013; Harmon 2003; Mendell 2006). As a matter of fact, contractors believe that the higher diesel fuel costs affect construction costs more considerably than higher petroleum-derived products (McFall 2005). This effect is more significant for earthmoving and highway contractors.

Long-term Impact of Crude Oil Price Fluctuations on Construction Cost Variations

This dissertation also shows that the impact of oil price on construction costs is long-term using empirical study (in comparison to the impact of construction market on construction costs). This long-term impact can be related to the change in construction business resulting from high and volatile energy-related costs. According to McFall (2005), higher energy-related costs affect how construction business is conducted because contractors need to deal with surcharges and contract variances.

Impact of Macroeconomic Condition on Construction Cost Variations

Leading Indicators

Although the existing literature lacks a rigorous empirical analysis quantifying the temporal relationship between macroeconomic condition and construction cost variations, several studies suggest the significant impact of macroeconomic condition on construction costs (Akintoye et al. 1998; Wong and Ng 2010). The results of the empirical studies in this dissertation also show that consumer price index and producer price index (representing economic conditions) are among the leading indicators of CCI and have a long-lasting impact on construction costs. This is not surprising since consumer price index and producer price index are the best indicators of general cost and price changes. The results of this dissertation also confirm that construction cost variations follow the general inflation represented by consumer price index. The results are consistent with the results achieved by Ng et al. (2000) and and Wong and Ng (2010).

Although the temporal relationship between construction costs and macroeconomic condition represented by indicators, such as consumer price index and producer price index has not been rigorously quantified, this relationship had been assumed and used. For example, some companies compare the historical values of construction cost index and broad price indices for the entire economy, such as consumer price index and producer price index to explain the movements of construction cost indices (Humphreys, K. K. 2004). Moreover, the U.S. Department of Housing and Urban Development (HUD) used national Consumer Price Index to update its construction cost (HUD 2005).

Lagging Indicators

Not all the macroeconomic variables are the leading indicators of construction cost variations. Some macroeconomic variables are simply lagging indicators of construction cost. For example, although there is evidence in the literature that unemployment rate is leading indicator of construction cost variations (Runeson 1988; Akintoye et al. 1998), the results of empirical studies in this dissertation showed that past values of unemployment rate is not useful to predict CCI. In other words, unemployment rate is not leading indicator of construction cost variations. These results are consistent with the results achieved by Wong and NG (2010). These results are also consistent with the fact that unemployment rate is one of the most popular lagging economic indicators (Singh 2013). In other words, unemployment rate follow patterns and trends of the economy that has already happened. The findings of this dissertation suggest that federals funds rate, prime lending rate, and Dow Jones Industrial Average do also lag construction cost variations.

Impact of Construction Market Conditions on Construction Cost Variations

Leading Indicators

Although the existing literature lacks a rigorous empirical analysis quantifying the temporal relationship between construction market conditions and construction cost variations, several studies suggest the impact of construction market conditions on construction costs (Runeson 1988; Fellows 1988; Taylor and Bowen 1987; Skitmore 1987).

The results of the empirical studies in this dissertation also show that the number of housing starts and the number of building permits (representing construction market conditions) are among the leading indicators of CCI and have an almost immediate impact on the construction costs. This is not surprising since these two factors represent the level of construction activities in the years to come. In other words, these variables represent construction market condition in the following years. Large number of building permits and housing starts indicates that the construction market will be active and more construction activities are expected in the future.

Lagging Indicators

Not all the construction market variables are the leading indicators of construction cost variations. For example, the results of this research show that past values of construction spending is not useful to predict CCI. In other words, construction spending is not leading indicator of construction cost variations. These results are consistent with the Sicotte and Glitman's findings. Sicotte and Glitman (2011) analyzed state-level data on construction costs and found that highway expenditures have the little impact on costs. The findings of this dissertation suggest that average weekly hours do also lag construction cost variations and is not among the leading indicators of construction cost variations.

Comparison of Accuracy of VEC Models for Forecasting ENR Construction Cost Index (CCI) with VEC Models for Forecasting National Highway Construction Cost Index (NHCCI)

Comparison of the predictability of the VEC models for forecasting CCI (presented in Chapter 4) with the predictability of the VEC models for forecasting CCI (presented in Chapter 6) shows that the latter models are more accurate. This higher level of accuracy for the VEC models for forecasting NHCCI than the VEC models for forecasting CCI can be explained by assessing the leading indicators used in the VEC models. The best VEC models created in Chapter 4 and Chapter 6 include crude oil price. The VEC model for forecasting NHCCI is more accurate because crude oil price is one of the main components of NHCCI (created in Chapter 6).

CHAPTER 8

CONCLUSION

Sixteen indicators are selected as prime candidates for leading indicators of construction cost variations based on the comprehensive literature review: prime loan rate, housing starts (number of housing starts), building permits (number of building permits), unemployment rate, consumer price index, producer price index, gross domestic product, GDP implicit price deflator, money supply, construction spending, federal funds rate, Dow Jones industrial average, crude oil price, employment level in construction, average hourly earnings, and average weekly hours. These variables depict the national macroeconomic conditions, energy cost, and construction market conditions.

Empirical tests are used to identify the leading indicators of CCI from the pool of candidate (potential) leading indicators representing the U.S. construction and economic environment. The Pearson correlation analysis is used as the preliminary analysis tool to test the relevance of the above potential leading indicators to CCI. Based on the correlation results and test statistics, consumer price index (+0.99), federal funds rate (-0.59), unemployment rate (-0.63), employment level in construction (+0.94), Average work hours (+0.71), prime loan rate (-0.49), money supply (+0.98), average hourly earnings (+0.99), Dow Jones industrial average (+0.92), crude oil price (+0.66), producer price index (+0.99), building permits (0.24), construction spending (+0.97), GDP (+0.99), and GDP implicit price deflator (+0.99) are significantly correlated with CCI at 1% significance level. These results support the selection of the above variables as prime candidates for leading indicators.

The results of Granger Causality tests show that consumer price index, crude oil price, and producer price index consistently Granger cause CCI at all the specified lag lengths. Therefore, consumer price index, crude oil price, and producer price index are consistent leading indicators of CCI. Building permits and housing starts Granger cause CCI in lower lag lengths. Hence, building permits and housing starts are among the leading indicators of CCI.

The results of Johansen's cointegration tests show that CCI and various combinations of the leading indicators are cointegrated. Hence, Vector Error Correction (VEC) models are created for multivariate time series modelling. The bivariate VEC model including CCI and crude oil price pass all diagnostic tests (Breusch–Godfrey serial correlation Lagrange multiplier test, ARCH test, and normality test). Two Bivariate VEC models (the VEC model including CCI and producer price index and the VEC model including CCI and crude oil price) provide better out-of-sample forecasts than the univariate time series models. Among bivariate VEC models, the model including CCI and producer price index provides the best out-of-sample forecasts.

The predictability of the multivariate time series modeling for forecasting CCI is tested using stochastically simulated data (Simulated CCI and crude oil price). Crude oil price is simulated using Geometric Brownian Motion (GBM). CCI is simulated using Gaussian Process that is considering the relationship between CCI and crude oil price over time. Multivariate and univariate time series models are created using the simulated data and the predictability of univariate and multivariate time series models are compared. The results show that the multivariate modeling is more accurate than univariate modeling for forecasting CCI.

Based on the Granger causality tests, crude oil price and average hourly earnings in the construction industry are selected as leading indicators of National Highway Construction Cost Index (NHCCI). In other words, past values of these variables contain information that is useful for forecasting NHCCI. Based on the results of cointegration tests, Vector Error Correction (VEC) models are created as the proper multivariate time series models to forecast NHCCI. Our results show that the VEC model including

NHCCI and crude oil price, and the VEC model including NHCCI, crude oil price, and average hourly earnings pass diagnostic tests. These VEC models are also more accurate than univariate models for forecasting NHCCI in terms of out-of-sample prediction error and out-of-sample mean square error.

Crude oil price was found as the only common leading indicator of both CCI and NHCCI. It was shown that crude oil price provides significantly useful information to predict the future values of CCI and NHCCI (Sicotte and Glitman 2011; Olatunji 2010; Damnjanovic and Zhou 2009; Mendell 2006; Gallagher and Riggs 2006; Reid 2005; Wilmot and Cheng 2003). The impact of fluctuations of oil price on construction costs can be supported by several observations in the construction industry. Contractors and suppliers are more cautious when bidding in a volatile oil market. Hence, they include higher contingencies in the projects. Therefore, the volatility of oil price propagates to the construction industry (Damnjanovic and Zhou, 2009). The cost of crude oil price affects directly the cost of some construction materials, such as asphalt cements since it is one of their main components (Akimovs, 2013; Damnjanovic and Zhou, 2009; Harmon 2003; Mendell 2006). The effects of volatile oil prices go beyond the price of asphalt materials by affecting the costs of gasoline and diesel fuel and consequently impacting the costs of construction operations that are all energy consuming (Davenport 2008; Akimovs, V. 2013; Harmon 2003; Mendell 2006). As a matter of fact, contractors believe that the higher diesel fuel costs affect construction costs more considerably than higher petroleum-derived products (McFall 2005). This effect is more significant for earthmoving and highway contractors. It was also shown that the impact of oil price on construction costs is long-term using empirical study (in comparison to the impact of construction market on construction costs). This long-term impact can be related to the change in construction business resulting from high and volatile energy-related costs. According to McFall (2005), higher energy-related costs affect how construction business is conducted because contractors need to deal with surcharges and contract variances.

Although the existing literature lacks a rigorous empirical analysis quantifying the temporal relationship between macroeconomic condition and construction cost variations, several studies suggest the significant impact of macroeconomic condition on construction costs (Akintoye et al. 1998; Wong and Ng 2010). The results of the empirical studies in this dissertation also show that consumer price index and producer price index (representing economic conditions) are among the leading indicators of CCI and have a long-lasting impact on construction costs. This is not surprising since consumer price index and producer price index are the best indicators of general cost and price changes. The results of this dissertation also confirm that construction cost variations follow the general inflation represented by consumer price index. Not all the macroeconomic variables are the leading indicators of construction cost variations. Some macroeconomic variables are simply lagging indicators of construction cost. For example, although there is evidence in the literature that unemployment rate is leading indicator of construction cost variations (Runeson 1988; Akintoye et al. 1998), the results of empirical studies in this dissertation showed that past values of unemployment rate is not useful to predict CCI. In other words, unemployment rate is not leading indicator of construction cost variations. My results are consistent with the results achieved by Wong and NG (2010). My results are also consistent with the fact that unemployment rate is one of the most popular lagging economic indicators (Singh 2013). In other words, unemployment rate follow patterns and trends of the economy that has already happened. The findings of this dissertation suggest that federals funds rate, prime lending rate, and Dow Jones Industrial Average do also lag construction cost variations.

Although the existing literature lacks a rigorous empirical analysis quantifying the temporal relationship between construction market conditions and construction cost variations, several studies suggest the significant impact of construction market conditions on construction costs (Runeson 1988; Fellows 1988; Taylor and Bowen 1987; Skitmore 1987). The results of the empirical studies in this dissertation also show that the number of housing starts and the number of building permits (representing construction market conditions) are among the leading indicators of CCI and have an almost immediate impact on construction costs. This is not surprising since these two factors represent the level of construction activities in the years to come. In other words, these variables represent construction market condition in the following years. Large number of building permits and housing starts indicates that the construction market will be active. Not all the construction market variables are the leading indicators of construction cost variations. For example, it was found that past values of construction spending is not useful to predict CCI. In other words, construction spending is not leading indicator of construction cost variations. These results are consistent with the Sicotte and Glitman's findings. Sicotte and Glitman (2011) analyzed state-level data on construction costs and found that highway expenditures have little impact on costs. The findings of this dissertation suggest that average weekly hours do also lag construction cost variations.

Contribution to the State of Knowledge

The leading indicators of CCI and NHCCI are rigorously identified. Multivariate time series models are created that are more accurate than the current univariate time series models for forecasting CCI and NHCCI. It was shown that crude oil price is significantly helpful for forecasting construction cost variations. It was shown that the macroeconomic variables, such as consumer price index and producer price index are valuable for forecasting medium- to long-term changes in construction costs. It was also shown that the construction market variables, such as the number of building permits and the number of housing starts do not have a long-lasting impact on construction costs. Based on this result, the cost estimator might not rely on the information about changes in the number of building permits and number of housing starts to forecast medium- to long-term changes in construction costs.

Contribution to the State of Practice

Appropriate construction cost forecasting models are created that enable cost estimators to take advantage of the contextual information for construction cost forecasting. This helps cost engineers and investment planners of capital and lengthy projects prepare more accurate bids, cost estimates, and budgets. The temporal relationship between construction cost variations and the macroeconomic, energy and construction market indicators in which the construction cost is changing are determined. This temporal relationship can be used to have a better understanding about the changes in construction costs in the U.S. economic and energy market context. For example, a cost estimator can continuously observe crude oil price and if he observes a pattern in changes of crude oil price, he can legitimately anticipate a pattern in changes in construction costs a few months down the road. This information has not been conventionally utilized for forecasting CCI.

Recommendations for Further Research

Several research paths forward can be imagined for this research. Here, some recommendations for further research are provided:

- This study was implemented for the U.S. economic, construction and energy market context. It could be interesting to go beyond the U.S. context. It is interesting to determine the temporal relationship between construction cost variations and global economic and energy indicators.
- This study was limited to the construction cost variations at the national level and spatial variations of costs in the U.S. are not determined. It could be interesting to characterize the construction cost variations in various states and regions in the U.S.

- The potential leading indicators in this study are representative of macroeconomic, energy and construction market at high level. For example, the employment level in construction shows the combined level of employment in all trades in the construction industry. It could be interesting to study lower level indicators. For example, it is recommended to test whether the employment level in a specific trade in the construction industry in a specific region is among the leading indicators of construction costs. This analysis might require data collection. For example, it might be required to collect data about specific construction trades or labor market condition in a specific region. This analysis at lower level might provide the opportunity to study state of the art issues in the construction market, such as craft availability and their impact on construction costs.
- There are rare shocks in the time series of CCI that cannot be modeled with the proposed models in this thesis. These shocks are the results of large jumps in CCI due to the large, sudden and unexpected price changes. Modeling these shocks is one of the recommendations for the future work.
- This study focused on crude oil price as the only indicator representing energy market. Several other energy indicators, such as natural gas price might also be among the leading indicators of CCI. Studying other energy related indicators, such as natural gas as the leading indicators of construction cost and price indices is highly recommended.
- Cost overrun is one of the most important risks that owner organizations face in the construction industry. It is recommended to conduct research to find how better forecasting of construction cost variations can help to reduce uncertainties about construction project costs and cost overruns. It is also recommended to conduct research to know how these cost overruns are related to the economic condition, and construction and energy markets.

- It is recommended to develop an index-based cost estimation approach that leverages the advancement in forecasting construction cost variations to improve the accuracy of project estimates.
- This study focused on a limited number of candidate leading indicators. There might be other candidate leading indicators to study. For example, it could be interesting to test whether the number of bidders is the leading indicator of construction cost indices (The number of bidders represents competition in the construction market). However, this analysis requires collection of data about the number of bidders at the national, regional, or project level and test whether the number of bidders is the leading indicator of construction costs at the national, regional, or project level.
- This study focused on cost indices representing heavy and highway construction. It is interesting to implement this study for other industry sectors, such as infrastructure, commercial, residential, and light industrial.

APPENDIX A

ALL MULTIVARIATE TIME SERIES MODELS THAT CAN BE CREATED WITH CCI AND ITS LEADING INDICATORS

Table A.1 show the out-of-sample predictability of the CCI univariate time series forecasting models. Table A.2 to A.6 show all the multivariate time series models that can be created with the CCI and its five identified leading indicators. These results show that there are several VEC models that are more accurate than the univariate time series models for forecasting CCI. VEC models with the following combinations of CCI and leading indicators provide better out-of-sample forecasts than existing univariate models:

- CCI, COP
- CCI, PPI
- CCI, HS, COP
- CCI, HS, PPI
- CCI, BP, COP
- CCI, BP, PPI
- CCI, COP, PPI

Although these results show that there are several VEC models that provide better forecasts than existing univariate models, the results of diagnosis tests (presented in Table A.7 to A.11) show that only the VEC model including CCI and COP pass diagnosis tests.

Table A.1: Out-of-sample predictability of the CCI univariate time series forecasting models

Model	MAPE	MSE
Seasonal ARIMA	1.40%	17921.6
HW-ES	2.68%	86890.7

Table A.2: Out-of-sample predictability of the CCI multivariate time series models with two variables

Model	MAPE	MSE
VECM (CCI, CPI)	1.48%	20594.0
VECM (CCI, HS)	2.26%	44304.61
VECM (CCI, BP)	2.41%	51450.38
VECM (CCI, COP)	0.96%	10544.9
VECM (CCI, PPI)	0.84%	7600.1

Table A.3: Out-of-sample predictability of the CCI multivariate time series models with three variables

Model	MAPE	MSE
VECM (CCI, CPI, HS)	2.37%	58969.36
VECM (CCI, CPI, BP)	2.71%	78297.12
VECM (CCI, CPI, COP)	2.99%	75580.79
VECM (CCI, CPI, PPI)	3.27%	100732.60
VECM (CCI, HS, BP)	3.04%	83207.70
VECM (CCI, HS, COP)	0.81%	7474.99
VECM (CCI, HS, PPI)	0.41%	1800.46
VECM (CCI, BP, COP)	0.92%	9926.88
VECM (CCI, BP, PPI)	0.50%	2920.29
VECM (CCI, COP, PPI)	0.94%	11246.75

Table A.4: Out-of-sample predictability of the CCI multivariate time series models with four variables

Model	MAPE	MSE
VECM (CCI, CPI, HS, BP)	3.09	103363.50
VECM (CCI, CPI, HS, COP)	4.27	163454.70
VECM (CCI, CPI, HS, PPI)	3.88	157969.6
VECM (CCI, CPI, BP, COP)	4.80	206552.1
VECM (CCI, CPI, BP, PPI)	4.24	188004.8
VECM (CCI, CPI, COP, PPI)	3.58	115313.6
VECM (CCI, HS, BP, COP)	1.10	13238.82
VECM (CCI, HS, BP, PPI)	0.55	3376.78
VECM (CCI, HS, COP, PPI)	2.05	38744.93
VECM (CCI, BP, COP, PPI)	2.43	54318.59

Table A.5: Out-of-sample predictability of the CCI multivariate time series models with five variables

Model	MAPE	MSE
VECM (CCI, CPI, HS, BP, COP)	4.93%	221511.0
VECM (CCI, CPI, HS, BP, PPI)	3.80%	148964.5
VECM (CCI, CPI, HS, COP, PPI)	4.51%	178669.3
VECM (CCI, CPI, BP, COP, PPI)	4.95%	215450.7
VECM (CCI, BP, COP, HS, PPI)	3.41%	109137.7

Table A.6: Out-of-sample predictability of the CCI multivariate time series model with six variables

Model	MAPE	MSE
VECM (CCI, CPI, HS, BP, COP, PPI)	4.93%	218736.9

Table A.7: Results of Breusch-Godfrey LM and ARCH tests for the residuals of VEC models with two variables

Model	Breusch-Godfrey test statistics	ARCH statistics
VECM (CCI, CPI)	42.51*	16.16*
VECM (CCI, HS)	157.60**	18.23
VECM (CCI, BP)	175.45**	16.50
VECM (CCI, COP)	30.25	11.82
VECM (CCI, PPI)	35.97	18.50**

*Notes: * indicates rejection of the null hypothesis at the 5% significance level; ** indicates rejection of the null hypothesis at the 1% significance level.*

Table A.8: Results of Breusch-Godfrey LM and ARCH tests for the residuals of VEC models with three variables

Model	Breusch-Godfrey test statistics	ARCH statistics
VECM (CCI, CPI, HS)	217.77**	18.24
VECM (CCI, CPI, BP)	234.52**	16.90
VECM (CCI, CPI, COP)	154.99**	12.24
VECM (CCI, CPI, PPI)	199.90**	14.36
VECM (CCI, HS, BP)	273.50**	16.54
VECM (CCI, HS, COP)	243.36**	19.87*
VECM (CCI, HS, PPI)	230.57**	15.26
VECM (CCI, BP, COP)	244.66**	707.36**
VECM (CCI, BP, PPI)	248.63**	13.68
VECM (CCI, COP, PPI)	155.16**	14.00

*Notes: * indicates rejection of the null hypothesis at the 5% significance level; ** indicates rejection of the null hypothesis at the 1% significance level.*

Table A.9: Results of Breusch-Godfrey LM and ARCH tests for the residuals of VEC models with four variables

Model	Breusch-Godfrey test statistics	ARCH statistics
VECM (CCI, CPI, HS, BP)	357.00**	18.95*
VECM (CCI, CPI, HS, COP)	370.91**	13.00
VECM (CCI, CPI, HS, PPI)	341.13**	16.08
VECM (CCI, CPI, BP, COP)	387.31**	14.46
VECM (CCI, CPI, BP, PPI)	358.69**	15.41
VECM (CCI, CPI, COP, PPI)	330.50**	11.23
VECM (CCI, HS, BP, COP)	378.99**	18.83
VECM (CCI, HS, BP, PPI)	370.82**	12.18
VECM (CCI, HS, COP, PPI)	373.63**	16.73
VECM (CCI, BP, COP, PPI)	400.20**	16.12

*Notes: * indicates rejection of the null hypothesis at the 5% significance level; ** indicates rejection of the null hypothesis at the 1% significance level.*

Table A.10: Results of Breusch-Godfrey LM and ARCH tests for the residuals of VEC models with five variables

Model	Breusch-Godfrey test statistics	ARCH statistics
VECM (CCI, CPI, HS, BP, COP)	514.05**	12.68
VECM (CCI, CPI, HS, BP, PPI)	499.98**	15.98
VECM (CCI, CPI, HS, COP, PPI)	496.79**	11.16
VECM (CCI, CPI, BP, COP, PPI)	519.46**	12.94
VECM (CCI, BP, COP, HS, PPI)	550.71**	17.40

*Notes: * indicates rejection of the null hypothesis at the 5% significance level; ** indicates rejection of the null hypothesis at the 1% significance level.*

Table A.11: Results of Breusch-Godfrey LM and ARCH tests for the residuals of VEC model with six variables

Model	Breusch-Godfrey test statistics	ARCH statistics
VECM (CCI, CPI, HS, BP, COP, PPI)	688.8**	11.4

*Notes: * indicates rejection of the null hypothesis at the 5% significance level; ** indicates rejection of the null hypothesis at the 1% significance level.*

APPENDIX B

DESCRIPTION OF THE CANDIDATE (POTENTIAL) LEADING INDICATORS

A pool of 16 candidate (potential) leading indicators is selected based on a comprehensive literature review about construction cost and price variations. Chapter 2 provides details about the comprehensive literature review resulted in selecting these factors. The availability of these indicators in the U.S. is considered in the selection of the potential leading indicators of the construction cost variations. These potential leading indicators are described here:

Prime Loan Rate (PLR)

Prime Loan Rate (PLR) is a rate applied by top-25 major U.S. insured commercial banks to price short-term business loans. This data is available monthly. (Data Source: Board of Governors of the Federal Reserve Systems)

Housing Starts (HS)

Housing Starts (HS) is the number of new privately owned housing units that their construction has been started in a given period. Excavation for foundation and footing is considered as the "start" of the construction process. This data is available monthly. (Data Source: U.S. Bureau of Census)

Building Permits (BP)

Building Permits (BP) is the number of new privately owned housing units that have been authorized to be constructed in a given period. This data is available monthly. (Data Source: U.S. Bureau of Census)

Unemployment Rate (UR)

Unemployment Rate (UR) is the percent of the U.S. labor force that is unemployed. This data is available monthly. (Data Source: U.S. Bureau of Labor Statistics)

Consumer Price Index (CPI)

Consumer Price Index (CPI) is measure of the price level of a representative basket of goods and services purchased by urban consumers. It is one of the widely used measures of inflation. This data is available monthly. (Data Source: U.S. Bureau of Labor Statistics)

Producer Price Index (PPI)

Producer Price Index (PPI) is “the average change over time in the selling prices received by domestic producers for their output”. This data is available monthly. (Data Source: U.S. Bureau of Labor Statistics)

Gross Domestic Product (GDP)

Gross Domestic Product (GDP) is a measure of the total value of goods and services that are produced in a country in a given period. It represents the economic health of a country. This data is available quarterly. (Data Source: U.S. Bureau of Economic Analysis)

Gross Domestic Product-Implicit Price Deflator (GDPIPD)

Gross Domestic Product-Implicit Price Deflator (GDPIPD) is a measure of the level of prices of all goods and services that are produced in a country in a given period. This measure is the ration of nominal GDP and real GDP. Nominal and real GDP are current-dollar and constant-dollar GDP. This measure accounts for inflation. This data is available quarterly. (Data Source: U.S. Bureau of Economic Analysis)

Money Supply (MS)

Money Supply (MS) is a type of measure that represents the amount of money in the nation's economy. In the United States, at least two types of money supply's measures are tracked: M1 and M2. M1 consists of currency in the public banks, institutions and the U.S treasury, traveler's checks, demand deposits, and other checkable deposits. M2 consists of M1 as well as savings deposits, time deposits less than $100, and balances in retail money market mutual funds. This data is available monthly. We used M2, which is a broader measure, in this research (Data Source: Board of Governors of the Federal Reserve Systems)

Construction Spending (CS)

Construction Spending (CS) is a measure of the value of new construction activities. It includes residential projects, non-residential projects, and public projects. This data is available monthly. (Data Source: U.S. Census Bureau)

Federal Funds Rate (FFR)

Federal Funds Rate (FFR) is the interest rate at which banks and other depository institutions charge each other for loans. This data is available monthly. (Data Source: Board of Governors of the Federal Reserve Systems)

Dow Jones Industrial Average (DJIA)

Dow Jones Industrial Average (DJIA) is an index representing the prices of shares of the 30 large U.S.-based companies. This data is available monthly. (Data Source: Yahoo Finance)

Crude Oil Price (COP)

Crude Oil Price (COP) is the domestic first purchase price of a barrel of crude oil. This data is available monthly. (Source: U.S. Energy Information Administration) Employment Level in Construction (ELC) is the number of employees (in thousands) on payrolls in construction. This data is available monthly. (Data Source: U.S. Bureau of Labor Statistics)

Average Hourly Earnings (AHE)

Average Hourly Earnings (AHE) is the average of hourly earnings in construction. This data is available monthly. (Data Source: U.S. Bureau of Labor Statistics)

Average Weekly Hours (AWH)

Average Weekly Hours (AWH) is average of weekly hours in the construction industry. This data is available monthly. (Data Source: U.S. Bureau of Labor Statistics)

REFERENCES

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

Akimovs, V. (2013). “Analysis of Road Construction Costs”, 28th International Baltic Road Conference, Vilnius, Lithuania, 26-28 August.

Akintoye, A., Bowen, P., and Hardcastle, C. (1998). “Macro-economic leading indicators of construction contract prices.” Construction Management and Economics, 16, 159-175.

Al-Tabtabai, H., Alex, A. P., and Tantash, M. (1999). “Preliminary cost estimation of highway construction using neural networks.” Cost Engineering, 41(3), 19–24.

Ashuri, B., and Lu, J. (2010). “Time series analysis of ENR construction cost index.” Journal of Construction Engineering and Management, ASCE, 136(11), 1227-1237.

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

Bell, L. C., & Bozai, G. A. (1987). “Preliminary cost estimating for highway construction projects.” AACE Transactions, C.6.1–C.6.4.

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

Chan, S.L. (2002). “Responses of selected economic indicators to construction output shocks: the case of Singapore.”, Construction management and Economics, 20, 523-533.

Chengalur-Smith, I. N., Ballou, D. P., and Pazer, H. L. (1997). “Modeling the costs of bridge rehabilitation.” Transportation Research Part A, 31(4), 281–293.

Chikobvu, D. (2013). "Random walk or mean reversion? Empirical evidence from crude oil market.“, İSTATİSTİK, JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION 6(1), pp. 1-9.

Davenport, B. (2008), “ACPA Concrete Pavement Progress”, American Concrete Pavement Association, Volume 44, Number 3.

Dayton, K. J. (2006). “Recent trends in highway construction costs.” 2006 WASHTO Subcommittee on Construction, Boise, Id.

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

Damnjanovic, I., and Zhou, X. (2009). "Impact of crude oil market behaviour on unit bid prices: the evidence from the highway construction sector." Construction Management and Economics, 27.9 (2009): 881-890.

Dancik, G. (2007). "mlegp: an r package for gaussian process modeling and sensitivity analysis.", http://ftp.uni-bayreuth.de/math/statlib/R/CRAN/doc/vignettes/mlegp/mlegp.pdf, last visit: 4/20/14.

Dancik, G.M., and Dorman, K.S. (2008). "mlegp: statistical analysis for computer models of biological systems using R.", Bioinformatics 24(17), 1966-1967.

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

ENR (2014). Engineering News Record, http://enr.construction.com/economics/, Last visit: April/20/2014.

ENR (2013). “100 Years of ENR Cost Indexes”, Engineering News Record.

Ervin, E. (2007). “How to protect profit as material prices rise.” Puget Sound Business Journal (Seattle), http://www.bizjournals.com/seattle/stories/2007/06/11/focus10.html?page=all

Fellows, R.F. (1988). “Escalation management.” Unpublished PhD Thesis, Department of Construction Management, University of Reading, UK.

FHWA (2012), “Request for Proposal: Index-based Cost Estimation with Accuracy and Precision Analysis”, <https://www.fbo.gov/index?s=opportunity&mode=form&tab=core&id=ecb700ff718bee50c575f8fe82f57dac>

FHWA (2014). Federal Highway Administration <http://www.fhwa.dot.gov/policyinformation/nhcci.cfm>, (3/30/2014)

Gallagher, J.P.(2008). “Material price escalation: Managing the risks.” Construction eNewsletter, http://www.troutmansanders.com/firm/media/mediadetail.aspx?media=703

Geman, H. (2007) "Mean reversion versus random walk in oil and natural gas prices.“, Advances in Mathematical finance, Birkhäuser Boston, pp. 219-228.

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. and 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.

Harmon, B. (2004). “Cost of Construction Materials Soaring”, Engineering Field Notes, Volume 36, Issue 1—2004.

Hegazy, T., and Ayed, A. (1998). “Neural network model for parametric cost estimation of highway projects.” Journal of Construction Engineering and Management, 124(3), 210–218.

HUD (2005). “Construction Cost Indices”, Prepared by NAHB Research Center, Inc. and Columbia Enterprises, Inc., Prepared for U.S. Department of Housing and Urban Development Office of Policy Development and Research Office of Housing, http://www.huduser.org/portal/publications/costindices.pdf, last visited: 4/5/2015.

Humphreys, K. K. (2004). “Project and Cost Engineers' Handbook”, CRC Press.

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.

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

Lean, C.S. (2001) Empirical tests to discern linkages between construction and other economic sectors in Singapore. Construction Engineering and Management, 19, 355-363.

McFall, K. 2005. “Crude Oil Prices Hit New Heights”, ENR Cost Report. 3Q: 71.

Mendell, B.C. (2006). “Benchmarking Construction Cost Changes in the U.S. South. American Reconstruction”, September 2006, http://www.forisk.com/UserFiles/File/Construction%20Index%20final.pdf.

NG, S.T., Cheung, S.O., Skitmore, R.M., Lam, K.C., and Wong, L.Y. (2000). “Prediction of tender price index directional changes.” Construction Management and Economics, 18, 843-852.

O’Haren, D.M.(2004). Material price escalation and availability, 24th Int. Risk Management Institute (IRMI), Orlando, Fla, <http://www.irmi.com/conferences/crc/handouts/crc24/mondaycrm/materialpriceescalationandavailability.pdf>

Osterwald-Lenum, M. (1992). “A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics.” Oxford Bulletin of Economics and Statistics, 54(3), 461-472.

Pewdum, W., Rujirayanyong, T., & Sooksatra, V. (2009). “Forecasting final budget and duration of highway construction projects.” Engineering, Construction and Architectural Management, 16(6), 544-557

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

Qian, P. Z., Wu, H., & Wu, C. J. (2008). “Gaussian process models for computer experiments with qualitative and quantitative factors.”, Technometrics, 50(3).

Reid, R.L. (2005). “Oil Prices, Hurricane Katrina Portend Higher Construction Costs, Civil Engineering, October, 24-25.

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

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

Saito, M., Sinha, K.C., and Anderson, V.L. (1991). “Statistical models for the estimation of bridge replacement costs.” Transportation Research Part A, 25A(6), 339–350.

Sanders, S.R., Maxwell, R.R., and Glagola, C.R. (1992). “Preliminary estimating models for infrastructure projects.” Cost Engineering, 34(8), 7–13.

Sicotte, R., and Glitman, K. (2011). “Is There a Link Between Highway Funding, Construction Costs and Employment?”, No. TRC Report# 10-012.

Singh, R. (2013). "Study of the Impact of Macro Economic Variable & Their Role as an Indicators for the S&P CNX Nifty.", A National Journal of Management, 1.1.

Skitmore, R.M. (1987). “Construction market: The market effect.” Royal Institution of Chartered Surveyors, Education Trust, Department of Civil Engineering, The University of Salford, Salford, UK.

Sthapit, N., and Mori, H. (1994). “Model to estimate highway earthwork cost in Nepal.” Journal of Transportation Engineering, 120(3), 498–504

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

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

Tse, R.Y.C., and Raftery, J. (2001) The effects of money supply on construction flows. Construction Management and Economics, 19, 9-17.

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

Williams, T. P. (2005). “Bidding ratios to predict highway project costs.” Engineering, Construction and Architectural Management, 12(1), 38–51.

Wilmot, C. G., and Cheng, G. (2003). “Estimating future highway construction costs.” Journal of Construction Engineering and Management, 129(3), 272–279.

Wilmot, C.G., and Mei, B. (2005). "Neural network modeling of highway construction costs." Journal of construction engineering and management, 131(7), 765-771.

Wong, J.M.W., Chiang, Y.H., and NG, T.S. (2008). “Construction and economic development: the case of Hong Kong.”, Construction Management and Economics, 26, 815-826.

Wong, J.M.W., and NG, S.T. (2010). “Forecasting construction tender price index in Hong Kong using vector error correction model.” Construction Engineering and Management, 28, 1255-1268.

VITA

SEYED MOHSEN SHAHANDASHTI

Seyed Mohsen Shahandashti is a researcher in the Economics of Sustainable Built Environment Lab at Georgia Institute of Technology (Georgia Tech), where he works under the direction of Dr. Baabak Ashuri. Besides pursuing PhD degree in construction, he also completed a master's degree in Operations Research at Georgia Tech. Prior to entering the PhD program, he completed a master's degree in Civil Engineering (Advanced Infrastructure Systems Track) from Carnegie Mellon University (CMU) and a master's degree in Civil Engineering from Nanyang Technological University (NTU).

In addition to the research presented in this dissertation, he had the opportunity of conducting research on several other projects as a graduate research assistant at Georgia Tech, CMU, and NTU. These research projects cover a variety of areas, such as Construction Project Portfolio Management, Sustainability and Building Energy Efficiency, Construction Productivity, Financial Valuation of Investments in Sustainable and Resilient Infrastructure Systems, and Economic Decision Analysis for High Performance Buildings.