VOLUME 31, NUMBER 5

December 2015

Abstract / Article
Page 2

References
Page 22

Author
Dr. Andrea Nana Ofori-Boadu

JTMAE The Journal of Technology, Management, and Applied Engineering

Exploring Regression Models for Forecasting Early Cost Estimates for High-Rise Buildings

Keywords:
Cost Estimating, Cost Forecasting, Cost Modeling, Regression Analysis, High-rise Buildings

PEER-REFEREED PAPER ■ PEDAGOGICAL PAPERS

ANDREA N. OFORI-BOADU is an Assistant Professor of Construction Management with the Department of Built Environment at North Carolina Agricultural and Technical State University. She earned her B.S. degree from the University of Science and Technology, Ghana (Building Technology, 1997). In 2004, she earned her M.S. (Industrial Technology) degree with a Construction Management specialization from North Carolina Agricultural and Technical State University. Her Ph.D. (Technology Management -Construction Management, 2012) was from Indiana State University. Her recent publications have been in Built Environment Project and Asset Management (BEPAM), International Journal of Engineering Research and Innovation (IJERI), Structural Survey, and the International Journal of Industrial Engineering & Production Research. Dr. Ofori-Boadu's research interests are in sustainability, quality management, and cost modeling. She is affiliated with the Association for the Advancement of Cost Engineering International (AACEI), the U.S. Green Building Council (USGBC), the Association of Technology Management and Applied Engineering (ATMAE) and the National Association of Women in Construction (NAWIC). Andrea is married to Victor Ofori-Boadu and they are blessed with three wonderful children. Dr. Ofori-Boadu may be reached at andreao@ncat.edu

Exploring Regression Models for Forecasting Early Cost Estimates for High-Rise Buildings

Dr. Andrea Nana Ofori-Boadu

ABSTRACT

Construction projects with inaccurate early cost estimates are burdened with excessive financial and contractual risks. Subsequently, construction investors and professionals demand accurate early cost estimates even when there are limited or no construction documents, because these estimates become the basis of project funding. Concerns that estimating accuracy during the early stages of construction projects lie between ±25% and ±50% present a dire need for more accurate forecasting models for specific building types. Forecasting models for early cost estimates for high-rise buildings are particularly of interest due to their complexity, high levels of investments and the paucity of related research. In order to determine the most accurate model for forecasting early costs for high-rise buildings in this present study, regression analysis methods were used to explore several cost drivers, cost estimating relationships and cost models.

The results revealed that the five key cost drivers are gross floor area, location, structural material, height, and completion date. With an improved R^2 of 64% and an error rate of approximately 9%, the most accurate model indicated that power functions best described cost estimating relationships. Also, the natural log of building cost per square foot was the most reliable dependent variable for modeling costs. This early cost forecasting model works best for high-rise buildings with gross floor areas ranging between 88,000 and 6,500,000 square feet.

Key results from this study are most useful to construction professionals and investors during the early phases of financial and economic feasibility decision making for high-rise building development. The proposed model has an improved level of accuracy, simplicity and should increase the ease of forecasting early cost estimates for high-rise projects, when there are limited or no construction contract documents. Academicians may use the processes outlined to educate their students on alternative early cost estimating methods. Future research should explore other methods and models for improving the accuracy of early cost estimates for specific building types. Improved cost forecasting models should reduce detrimental construction project results such as losses, delays, strife, and litigation. Accurate early cost forecasting models should improve overall construction project success and client satisfaction.

Introduction

Accurate early construction cost estimates are important for key economic decision making during the early phases of building design and development. Kim, An and Kang (2004) emphasized that the accurate estimation of construction project costs is critical to project success. In particular, the overall success of high-rise construction projects depends heavily on accurate early cost estimates because these estimates provide the basis for project funding and are essential for feasibility studies, project planning, implementation, and cost control strategies.

Typically, construction cost estimates are more accurate when construction contract drawings and specifications are 100% complete. As such it is very challenging to obtain accurate early estimates for high-rise buildings since the early phases of construction projects are commonly characterized with limited or no construction documents (Kim et al., 2004; Ofori-Boadu & Waller, 2014). Dagostino and Peterson (2011) concurred that there are no or very few detailed drawings and only vague descriptions of the function, performance or physical characteristics of projects exist. Without detailed drawings, the likelihood of obtaining inaccurate estimates is very high and both owners and contractors are risk averse, with low tolerance for overestimates and underestimates (Ahuja, Dozzi, & Abourisk, 1944; Cheun, Wong, & Skitmore, 2008). Generally, estimating inaccuracies increase the potential for cost overruns, losses, less than expected profits, delays failures, litigation, client dissatisfaction and other undesirable results.

In order to minimize the inaccuracies of early cost estimates, researchers and practitioners have developed several early cost estimating models and methodologies. Common methods include the use of square foot costs, historic data, cost indices, cost modeling, as well as expert and other statistical analysis (Hwang, 2009; Kim et al., 2004; Dagostino & Peterson, 2011). Statistical methods and models have been used to explain the cost estimating relationships (CERs) that exist between building costs and their cost drivers. Kwak and Watson (2005) defined a cost estimating relationship as a proven relationship between a certain characteristic and an estimated cost. In particular, Skitmore and Marston (1999) explained that cost models play an important role in cost estimating because when using cost models, perceived characteristics of a construction project are used to generate project cost estimates. The attractiveness of cost models lie in the fact the models are used to provide simple and straightforward answers to some of the very difficult and complex problems (Skitmore & Marston, 1999). Raftery (1984) listed the criteria for assessing the performance of building cost models to include data, data/model interface, model techniques, and interpretation of output. With the objective of generating debate on the building cost models, Skitmore and Marston (1999) raised concerns regarding the lack of expertise and sufficient research to advance the implementation of building cost modeling.

Review of Literature

A cost model is a set of mathematical equations that convert resource data to cost data (Business Dictionary, 2015). Building costs are functions of several categories of variables including building characteristics, socio-economic conditions, project characteristics and estimators' experience (Kim et al., 2004; Mahamid, 2011; Ofori-Boadu & Waller, 2014). Cost functions are used to demonstrate the mathematical formula that expresses the cost estimating relationship existing between building costs and their statistically significant independent variables (cost drivers).

Based on an assumption of linear functions, Kouskoulas and Koehn (1974) identified significant cost drivers to include location, time of realization, function, height, quality and technology. Karshenas (1984) concluded that floor areas and heights impacted adjusted building costs. Despite its limitations, McGarrity's parametric model revealed that gross floor area, number of stories, length of contract, liquidated damages and height had strong relationships with building costs. Also, the model testing results indicated that percent residuals were between 5.15% and 11.83% (McGarrity, 1988). In his regression analysis, Mahamid (2011) obtained R2 that ranged from 0.92 to 0.98 with the mean absolute percentage error ranging between 13.3% and 31%; while Sonmez (2008) obtained R2 values ranging between 0.75 and 0.96. The mean absolute percent error for Sonmez's study was 12% and the variables that were identified as significant included parking, site area, steel and concrete frame, gross building area and city cost index. Lowe, Emsley, and Harding (2006) revealed that their key cost drivers were gross internal floor area, function, duration, mechanical installations and piling. Trost and Oberlender (2003) concluded that the factors for predicting estimate accuracy included basic process design, team experience and cost information, time allowed to prepare the estimate, site requirements and bidding and labor climate. Lowe et al. (2006) proposed a log of cost backward model with an R2 of 0.661, and a mean absolute percentage error of 19.3%. Further, Lowe et al. (2006) concluded that traditional methods of cost estimation typically have mean absolute percentage errors (MAPE) of approximately 25%.

While in recent times the accuracy of early cost estimates have improved, Hwang (2009) emphasized that there is the need to improve the predictive capability of the construction industry. There is a critical need to explore early cost forecasting models which have lower percentage errors in order to improve the accuracy of early cost estimates. While several cost estimating models have been developed, the use of the traditional and older methods still outweigh the use of the newer models; probably because practitioners are not well-equipped to use the more elaborate models (Fortune & Lee, 1996). Cheung, Franko and Skitmore (2006) concurred that there is a need to develop alternative models, which are based on a logical and systematic method and are superior to the existing models.

Despite the fact that there have been some early estimating models developed in the past, no cost forecasting model was found to focus on high-rise buildings in the global context using recent cost data. Barr (2007) emphasized that high-rise buildings are important buildings to study in their own right since they are qualitatively different from most typical structures, and require vast amounts of financial capital and huge consortiums of investors, including banks and insurance companies. In so many different ways, building design and construction teams are faced with unique challenges when managing these high-rise projects which have scopes and complexities of work that are extensively beyond those of regularly sized buildings.

Also, high-rise building costs are exacerbated by additional costs commonly related to land acquisition; increased number of elevators and parking spaces; more stringent environmental, safety and building codes; insurance and bonding requirements; and other political, social and economic risks.

Background of the Problem

Construction investors and professionals heavily rely on early cost estimates to make decisions regarding the economic feasibility of construction projects. In order to minimize poor financial and economic feasibility decision making during the early phases of construction project development, there is a critical need for early cost estimates to be as close as possible to actual building costs (Kouskoulas & Koehn, 1974). This challenging goal is further complicated because of the absence of complete construction contract documents during the early phases of project development, wide variations among construction projects, and imperfect cost estimating methods (Kim et al., 2004). According Mahmid (2011), many difficulties arise during the estimation process due to the lack of preliminary information, lack of cost databases and lack of appropriate cost estimation methods.

While many researchers have focused on cost estimation methods and models that are generated from complete construction contract documents, there is a dire need for cost models research studies that accurately forecast early cost estimates for high-rise buildings using very limited or no construction contract data. Barr (2009) emphasized that surprisingly little work had been done on the economics of skyscrapers and "skynomics" remains relatively unexplored, despite the continued fascination by the public, journalists and scholars (p. 27). Skitmore and Marston (1999) emphasized that there is a considerable lack of knowledge in building cost modeling and suggested that modeling techniques could be the key to better understanding building costs.

Existing cost estimating methods are not 100% accurate (Kim et al., 2004). Mahamid and Amund (2010) found that 100% of the construction projects under study suffered from cost diverge; 76.33% were underestimated and 23.67% were overestimated. Among the projects studied, it was observed that 62% were overestimated and the rest were underestimated (ROad Costs Knowledge System, 2002). Estimating accuracy during the early stages of construction projects lies between ±25% and ±50% (Lowe et al., 2006; Schexnayder & Mayo, 2003). Early cost estimates based on inaccurate methods have been detrimental to construction project teams and resulted in losses, delays, strife, litigation and customer dissatisfaction (Ahuja et al., 1944; Assaf & Al-Hejji, 2006; Ofori-Boadu & Waller, 2014). Recent work in cost estimation indicates that there are still many problems in cost estimation at the conceptual stage of a construction project cycle (Mahamid, 2011). Given the importance of accurate estimates to the success of construction projects, additional research should be conducted to improve existing cost estimating methods and models.

Purpose of the Study

The purpose of this quantitative study was to develop a cost estimating model for forecasting early building cost estimates by exploring the cost estimating relationships that exist between high-rise building cost estimates and a set of cost drivers. The null hypothesis stated that no statistically significant relationships exist between early building cost estimates and the specified cost drivers. In particular, linear and non-linear cost estimating relationships that exist between cost drivers and building costs were assessed using multiple regression analysis (MRA).

MRA has the capacity to reveal statistically significant relationships that exist between a dependent variable and one or more independent variables. MRA models are generally expressed algebraically as shown in Equation 1:

$$Y = \boxtimes_0 + \boxtimes_1 X_1 + \boxtimes_2 X_2 + \quad + \boxtimes \qquad \text{Eq. (1)}$$

where:

Y	-	Dependent Variable
$\boxtimes_0$		Intercept or Constant (The expected Y value when X=0)
$\boxtimes_1, \boxtimes_2$	-	Regression Coefficients (The estimated slope of the regression line: it represents the expected change in Y for a unit change in X, controlling for all other Xs)
X_1, X_2	-	Independent Variables
$\boxtimes$	-	Random Errors

Significance of Study

This present quantitative research study is useful because the development of more efficient cost forecasting models could improve the accuracy of early cost estimates for high-rise buildings; consequently, enhancing the accuracy and reliability of financial and economic feasibility decisions. Accurate early cost forecasting models are particularly beneficial during the decision making phase because early stage building design decisions are more cost sensitive compared to decisions made later in the building process (Skitmore & Marston, 1999). Also, accurate estimates improve project planning, implementation and cost control strategies. Further, construction investors and professionals are provided a simple, quick, inexpensive and accurate model for predicting early costs for high-rise buildings with limited or no construction contract data. The model should minimize uncertainty, financial risks, losses, delays, litigation, and customer dissatisfaction, and other detrimental effects from inaccurate estimates. In the long term, accurate early cost estimating models should improve overall project success and customer satisfaction.

Limitation of Study

Since secondary data obtained from multiple sources were used in this study, any inaccuracies in the collection of the original data could impact the results from this study. Also, although productivity influences building costs, this variable was not investigated in this study. While an early cost estimating model was generated from this research study, the utilization of this model is limited to only high-rise buildings.

Methodology

This quantitative study relied on secondary data to develop adjusted building costs for 118 high-rise buildings constructed in over 20 different countries. Eight regression models were assessed to determine the best model for forecasting early cost estimates using limited data during the early phases of high-rise building project development. The high-rise buildings selected for this research study met the Council of Tall Buildings and Habitats (CTBUH) threshold of 165 feet or 14 stories (CTBUH, 2014).

Data Collection and Normalization

Secondary research data was primarily obtained from the tallest 500 buildings list published by the Council of Tall Buildings and Urban Habitats (CTBUH, 2014). Additional data on these high-rise buildings and their business environments were obtained from other published datasets available on the websites of the following organizations: Emporis; International Labor Organization (ILO); International Monetary Fund (IMF); U.S. Department of Labor (Bureau of Labor Statistics); and the World Bank. Since these high-rise buildings have different completion dates and currencies, their building costs and monthly labor wage rates were adjusted to 2005 real U.S. dollars using the Consumer Price Index (CPI) and time-specific foreign exchange rates.

Model Development

In this present study, eight different conceptual models (M1 – M8) were developed by varying the operationalization of each model's dependent variable (DV). Absolute cost is used as a dependent variable in two of these conceptual models because the ultimate goal of any estimate is to eventually obtain the absolute overall cost of the building. As such the relationships between cost drivers and absolute costs are still of interest to practitioners and researchers, and cannot be totally ignored. As such, the dependent variables for the two models were operationalized as building cost (BC) and the natural log of building cost (lnBC).

The limitations associated with models that rely on dependent variables with absolute values were compensated by introducing the six additional conceptual models which had dependent variables that had been normalized. The dependent variables for these six models were operationalized as follows: building cost per square foot (BCSF); building cost per story (BCST); and building cost per lineal feet of building height (BCLF); natural log of building cost per square foot (lnBCSF); natural log of building cost per story (lnBCST); and natural log of building cost per lineal feet of building height (lnBCLF).

According to Skitmore and Marston (1999), factors related to procedures, production and procurement have a major impact on the building cost. Also, building characteristics, socio-economic conditions, and project characteristics drive building costs and should be seriously considered when modeling building costs (Kim et al., 2004; Mahamid, 2011; Ofori-Boadu & Waller, 2014). These factors drive building costs and were adopted as the independent variables in this study. Based on existing literature and data availability, the 12 independent variables (IVs) initially selected for this study were as follows: Location (L); Ease of Doing Business (EB); Monthly Wages (MW); Completion Date (CD); Building Duration (BD); Type of Project (TP); Shape Complexity (SC); Height of Building (H); Number of Stories (ST); Height per Story (HS); Floor Area (FA); and Structural Material (SM) (Kouskoulas & Koehn, 1974; Karshenas, 1984; Lowe et al., 2006; Trost & Oberlender, 2003; Sonmez, 2008).

Models M1 to M4 assumed that linear cost estimating relationships existed between IVs and DVs, while non-linear cost estimating relationships were assumed for Models M5 to M8. Table 1 best captured the different variables used for M1 to M8.

TABLE (1): DVs and IVs for Models M1 to M8

Variables / Models	Category	M1	M2	M3	M4	M5	M6	M7	M8
DV	PC	BC	BCSF	BCST	BCLF	lnBC	lnBCSF	lnBCST	lnBCLF
L:NA=0; NNA.=1	SE	L	L	L	L	L	L	L	L
EB	SE	EB	EB	EB	EB	lnEB	lnEB	lnEB	lnEB
MW: 2005 U.S. $	SE	MW	MW	MW	MW	lnMW	lnMW	lnMW	lnMW
CD	PC	CD	CD	CD	CD	CD	CD	CD	CD
BD: years	PC	BD	BD	BD	BD	lnBD	lnBD	lnBD	lnBD
TP: Non-residential = 0; Residential = 1	PC	TP	TP	TP	TP	TP	TP	TP	TP
SC: Rectangular = 0 Non-rectangular = 1	BC	SC	SC	SC	SC	SC	SC	SC	SC
H: feet	BC	H	H	H	H	lnH	lnH	lnH	lnH
ST	BC	ST	ST	ST	ST	lnST	lnST	lnST	lnST
HS: feet / story	BC	HS	HS	HS	HS	lnHS	lnHS	lnHS	lnHS
FA: square feet	BC	FA	FA	FA	FA	lnFA	lnFA	lnFA	lnFA
SM: Steel only = 0; Concrete and steel = 1	BC	SM	SM	SM	SM	SM	SM	SM	SM

Results and Discussions

Out of the original data set of 500 high-rise buildings, only 118 were retained for the preliminary multiple regression analysis because the 382 buildings with incomplete data were all dropped from the study. Using the backward method for multiple regression analysis, the variables that were not statistically significant at $\alpha = 0.05$ were eliminated leaving only nine variables. The eliminated variables were BD, TP and SC, implying that building duration, building types and building shapes did not significantly impact high-rise building costs. Also, since the correlation matrix indicated that high levels of multicollinearity existed between H and ST, ST was eliminated to minimize biases and errors that could result from multicollinearity issues. Since the Q-Q plot exposed several outliers which could impose some extreme influences on the preliminary regression lines, 16 outliers were removed from the original dataset of 118, leaving 102 buildings for the final regression analysis. Further, two of the 102 buildings were reserved and used for testing the validity of the selected model, leaving 100 buildings for model development.

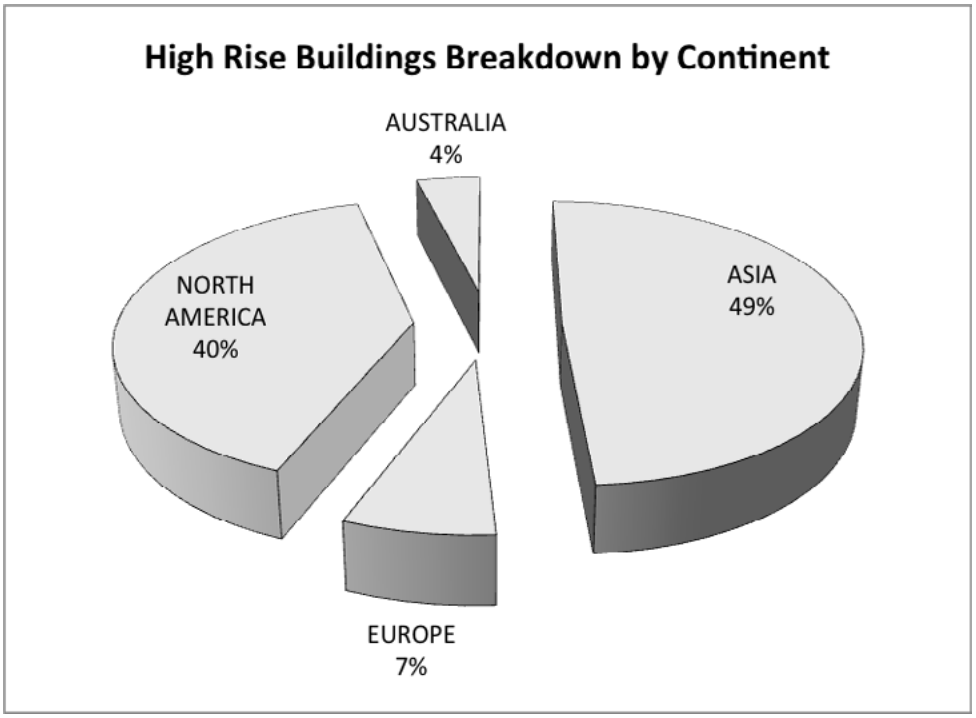
A final regression analysis was implemented using the eight original conceptual models, but with only eight IVs (L, EB, MW, CD, H, HS, FA, and SM). In order to better understand the characteristics of the high-rise buildings utilized for the final model development of Models M1 to M8, the results from the descriptive statistical analysis are shown in Table 2. Notably, the physical characteristics of these high-rise buildings such as the height and floor area are extremely high as these are some of the tallest buildings in the whole world. Consequently, the building costs are also very high as demonstrated in Table 2.

TABLE (2): Descriptive Statistics

Variable	Mean
Height	977 feet
Floor area	1,893,232 square feet
Height per story	14.19 feet
Monthly wage rate	$2,370.41
Ease of doing business	29.45
Completion Date	1997
Building cost	$384,322,255.00
Building cost per square foot	$252.32
Building cost per story	$5,503,070.00
Building cost per linear foot in height	$388,136.88

As indicated by the descriptive statistics, the dataset comprised of some of the tallest buildings in the world, with the average high-rise building having 69 stories. These high-rise buildings are considered as 'extremely tall' as they exceeded the CTBUH threshold of 14 stories. Forty-nine percent of the buildings are located in Asia, closely followed by 40% in North America as shown in Figure 1.

FIGURE (1): High Rise Buildings Breakdown by Continent



The United States of America had the most high-rise buildings (36%), followed by China with 13% and the United Arab Emirates (11%). The top tallest ten buildings in this data set are in Asia and included buildings such as the Shanghai World Financial Center and the Petronas Tower. Obviously, since these buildings are symbols or statements of power and technological advancements, most of them are located in the wealthier nations of the world.

As expected, by dropping the three IVs and removing the 16 outliers from the data used for the preliminary analysis, the R2 values of all the final models increased by an average of 0.12 as shown in Table 3. The outliers that were dropped from the dataset after the preliminary analysis were the buildings with extremely low and extremely high gross floor areas and building heights. The improvement in the models after the removal of the outliers suggested that these cost estimating models are most accurate when predicting building costs of high-rise buildings with gross floor areas ranging between 88,000 square feet and 6,500,000 square feet.

TABLE (3): Summary Results for Final Models

	M1	M2	M3	M4	M5	M6	M7	M8
Preliminary R^2	0.49	0.12	0.37	0.32	0.46	0.57	0.39	0.36
Final R^2	0.58	0.41	0.44	0.38	0.60	0.64	0.52	0.50
Change in R^2 (Final - Preliminary)	0.09	0.29	0.07	0.06	0.14	0.07	0.13	0.14
Number of statistically significant variables	5	5	6	5	6	6	7	6
Statistically significant variables	L,CD, H,FA, SM	L,CD, H,FA, SM	L,CD, H,FA, SM, HS	L,CD, H,FA, SM	L,CD, H,FA, SM, MW	L,CD, H,FA, SM, MW	L,CD, H,FA, SM, MW,HS	L,CD, H,FA, SM, MW

Consistent with the preliminary analysis, the best model fit for this data was Model M6 with its R2 of 0.64. As expected, the non-linear model with a dependent variable based on relative building costs per square foot of gross floor area continues to be the most reliable dependent variable for forecasting building costs. It is not uncommon to find estimators forecast construction costs based on unit gross floor area, as it consistently provides a reliable quantitative measure of the resources needed to construct buildings. In conformance with findings from the preliminary analysis, Model M6 used the natural logarithm of building cost per square foot reiterating that the power function is the preferred function when predicting building costs. This is a deviation from the linear relationship assumption, and could largely be due to the economies of scale. As the gross floor area of buildings increase, their associated building costs increase also, but at a reducing rate and thus the building cost per square foot reduces. A plausible explanation could stem from the fact that fixed costs such as overhead and land costs do not necessarily increase proportionally with building size.

Considering the limited quantitative data and project information available during the early stages of project development, the R2 value obtained seemed adequate. The implication is that the variables shown in Table 4 explained 64% of the relationship between the IVs and the natural log of building cost per square foot of gross floor area. The high levels of variation and risk that exist in the complex and dynamic global construction market render an R2 of 64% acceptable, especially compared to the R2 of 66% obtained by Lowe et al. (2006) which focused only on projects located in the United Kingdom. The obtained R2 of 0.64 indicates that 36% of the variation in the R2 could not be accounted for by the variables in this model, possibly because the data collected does not perfectly represent all the different IVs that could affect building costs for high-rise buildings in the global construction market. Further research studies should be conducted to explore other building cost drivers that are not investigated in this study. Also, further studies should focus on the use of other functions and statistical methods for accurately expressing the cost estimating relationship between high-rise building costs and their cost drivers. The regression coefficients for the statistically significant variables in Model M6 are shown in Table 4.

TABLE (4): Regression Coefficients for Model M6

Variable	Parameter Estimate	Standard Error	F Value	Pr > **F**
LOCATION	-0.68387	0.10498	42.44	<.0001
Ln MONTHLY WAGES	-0.06593	0.02913	5.12	0.0259
COMPLETION DATE	0.00694	0.00293	5.61	0.0199
Ln HEIGHT	1.62705	0.21913	55.13	<.0001
Ln FLOOR AREA	-0.65645	0.06635	97.88	<.0001
STRUCTURAL MATERIAL	-0.46513	0.11363	16.76	<.0001

The results shown in Table 4 are expressed in formula shown in Equation 2 below:

$$LnBCSF = -9.19 - 0.68\,(L) - 0.07\,(LnMW) + 0.007\,(CD) + 1.63\,(LnH) - 0.66\,(LnFA) - 0.47\,(SM) \quad \text{Eq. (2)}$$

where:

LnBCSF	Natural Log of Building Cost per Square Foot
L	Location
Ln MW -	Natural Log of Monthly Wages
CD	Completion Date
LnH	Natural Log of Building Height
LnFA	Natural Log of Gross Floor Area
SM	Structural Material Type

With a negative coefficient, L was statistically significant in all the eight models. Considering the fact that L was operationalized as a dummy variable for North American and non-North American high-rise buildings, the negative relationship confirmed that non-North America buildings were cheaper. The North American buildings are predominantly located in the United States, thus magnifying the effect from the U.S. socioeconomic environment. While the highly regulated U.S. business environment improves business efficiency, it is associated with higher material, labor, equipment and overhead costs. Further, U.S. high-rise building costs are burdened with additional expenses from highly unionized labor organizations; quality, safety and environmental standards; financial demands from stakeholders; and other political, economic and social issues.

MW had a negative coefficient and was statistically significant in four of the non-linear models. While MW was expected to have positive coefficients, the results revealed that negative relationships exist between MW and building costs – indicating that as MW increases, there are reductions in building costs. A conceivable explanation could be drawn from the theories associated with improved productivity. It is reasonable to assume that higher wages are paid to more highly trained and productive workers. As such, the countries with the higher average monthly wages could have higher levels of productivity which could result in improved efficiency. Also, highly skilled and experienced workers could also have increased access to technologically advanced equipment and resources leading to significant improvements in productivity. The gains in productivity levels arising from highly efficient and motivated workers with higher wages and more highly automated construction processes could reduce overall construction time, waste, errors, and eventually reduce building costs.

As expected, with a positive coefficient and a demonstration of statistical significance in all eight models, the results confirmed that CD increases building costs. Since all the high-rise building costs were normalized to 2005 U.S. dollars, this variable was used as an indicator of technological advancement and not necessarily as a measure of inflation. Due to the higher technological and aesthetic characteristics of high-rise buildings in recent times, newer buildings are more expensive compared to older buildings.

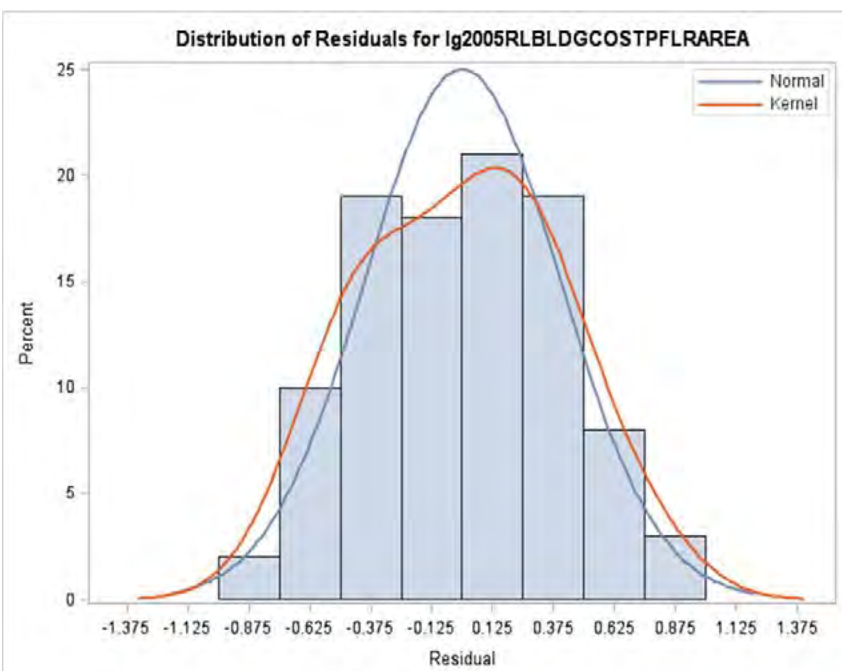
With its positive coefficient, H was statistically significant in all of the models. In support of the research hypothesis, the positive coefficient established the fact that building costs increase as the height of buildings increase - making H one of the most reliable variables for forecasting high-rise building costs. Taller buildings have higher costs estimates primarily due to the increased quantities related to materials, labor, equipment, and other resources. For example, the exterior cladding of high-rise buildings could be one of the more expensive cost items due to the need to provide these buildings with strong, weather resistive claddings and impressive aesthetics that are pleasing to the public. These claddings are directly related to the height and surface area of the buildings and as such taller buildings have significantly higher costs. This result also show that increasing building heights result in increases in building costs that cannot be significantly reduced by the cost savings from the economies of scale and learning curves as noted in FA.

FA in Model M6 had a negative coefficient and was statistically significant in all the eight models. This implies that although absolute building costs increased with FA, building costs per square foot reduced with FA. As previously explained, the decrease in building costs per square foot could be associated with the economies of scale drawing from the fact that fixed development costs do not necessarily always increase with gross floor area. Also the improvement in the learning curves related to repetitive work commonly associated with high-rise buildings, could increase productivity and efficiency, eventually leading to significant cost savings. FA continues to be the most reliable driver of building costs, as demonstrated by both the preliminary and final regression analysis.

The negative coefficient for SM implies that structural concrete framed buildings are cheaper than structural steel framed buildings. The improvements in concrete technology have increased the quality and strength of concrete, as well as the ease of working with concrete in high-rise buildings – making reinforced concrete a cheaper and viable option to structural steel. Also, the processing and transportation of structural steel requires a relatively higher amount of financial investment into plant, equipment and freight, adding on to its material costs.

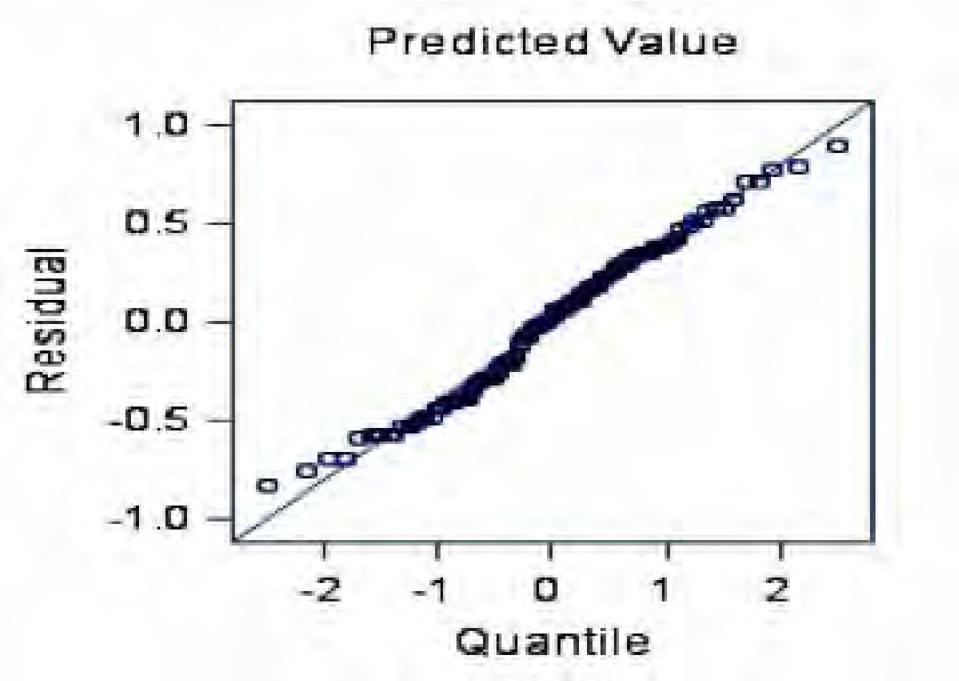
The statistical plots associated with the fit diagnostics for Model M6 were generated during statistical analysis of the data. In conformance to one of the major assumptions of multiple regression analysis, the histogram and normal curve confirmed that the residuals approximately followed the normal distribution as shown in Figure 2.

FIGURE (2): Distribution of Residuals



The quantile plot shown in Figure 3 revealed that with the exception of a few extreme points, most of the residual data in the quantile plot were close to or on the regression line, indicating a good fit of the model.

FIGURE (3): Quantile Plot



Two high-rise buildings located in the United States of America and in Asia were used to test the validity of Model M6. Building 1 which is located in the U.S. has a larger gross floor area and is taller than Building 2, which is located in Asia. Both buildings have a reinforced concrete structural frame and were completed two years apart of each other. The characteristics of these two buildings, as well as the results from the validity tests are summarized in Table 5. Further, Table 5 also showed that the building costs were higher for Building 1 when compared to Building 2.

TABLE (5): Validity Test Results for Model M6

Variables	Building 1	Building 2
Location	U.S.	Asia
Completion Date	2009	2007
Height	1389'	833'
Floor Area	2,600,001 sf	959,266 sf
Structural Material	Concrete	Concrete
Monthly labor wages	$3522.60	$3700.27
Observed cost (O)	$297.20 / sq. ft.	$146.99 / sq. ft.
Predicted cost (P)	$338.47 / sq. ft.	$140.63 / sq. ft.
Residuals, R = (P-O)	$47.27 / sq. ft.	-$6.36 / sq. ft
Percent residuals, PR = R /O*100	13.89%	-4.33%

From the model testing results, the PR for Buildings 1 and 2 were 13.89% and - 4.33%, respectively. Comparing the PRs for Buildings 1 and 2, it was concluded from Table 5 that Model M6 produces more accurate building costs per square foot for "smaller" high-rise buildings. Also, this model turns to underestimate the cost of smaller projects and overestimate the cost of larger projects. Following Kim et al., (2004), the Mean Absolute Error Rate (MAER) was used to measure the accuracy of Model M6. It was calculated using Equation 3 below:

$$MAER\ (\%) = \left(\Sigma \frac{|P-O|}{O}\right) * \frac{100}{n} \qquad \text{Eq. (3)}$$

Where n is the number of data sets

The MAER for the test data is 9.11%. Despite the wide variations among building costs at the global level and building heights in this dataset, the results obtained from Model M6 indicates that this model has reasonable levels of accuracy that are comparable to those obtained by McGarrity (1998) and Lowe et al. (2006).

Conclusion

The accuracy of early construction cost estimates is critical for early decision making regarding the economic feasibility of high-rise building projects, as well as overall project success. Using multiple regression analysis, early cost estimating models for forecasting high-rise building costs were explored in this research study. The five key cost drivers that appear significant in all the regression models were gross floor area, location, structural material, height, and completion date. With an R^2 of 0.64, the best model utilized the natural logarithm of building cost per square foot of gross floor area as its dependent variable. In addition to the previously noted cost drivers, this model had monthly labor wage rates demonstrating statistical significance. The test data showed that the mean absolute error rate for this model was approximately 9%, and that more accurate predictions were associated with smaller buildings. The model performed best for high-rise buildings with gross floor areas ranging between 88,000 square feet and 6,500,000 square feet.

The utilization of this model is limited to the forecasting of early cost estimates for high-rise buildings. A limitation of this study is that the recommended cost forecasting model did not assess the impact of productivity on building cost estimates. As such, future studies should assess other cost drivers and explore other methods for developing even more accurate models for forecasting early cost estimates for different building types.

The key results from this study provide construction investors and professionals with a cost estimating model that may be used to forecast early cost estimates for high-rise buildings at a reasonable level of accuracy. In addition, the regression methodologies demonstrated in this study may be used by design and construction professionals to develop their own cost forecasting models using historic construction cost data. The accurate prediction of construction cost estimates during the very early stages of project development should enhance overall project success, as it augments early and effective decision making that improve the planning, budgeting, financing, and overall management of high-rise building projects. Accurate estimates should improve cost control, project success and client satisfaction.

References

Ahuja, H., Dozzi, S., & Abourisk, S. (1944). *Project Management Techniques in Planning and Controlling Construction Projects*. New York: John Wiley and Sons, Inc.

Ali, M. M. (2001). Evolution of concrete skyscrapers: from Ingalls to Jin mao. *Electronic Journal of Structural Engineering, 1*(1), 2 – 14.

Assaf, S. A. & Al-Hejji, S. (2006). Causes of delay in large construction projects. *Journal of Construction Engineering and Management, 24*, 349 – 357.

Barr, J. (2012). Skyscraper height. *Journal of Real Estate Finance and Economics, 45*(3), 723 – 753.

Business Dictionary (2015). *Definitions - Cost Model*. Retrieved from http://www.businessdictionary.com/definition/cost-model.html

Chan, A. P. C., Scott, D., & Chan, A. P. L. (2004). Factors affecting the success of a construction project. *Journal of Construction Engineering and Management, 130*, 153 – 155.

Chan, S. L. & Park, M. (2005). Project cost estimation using principal component regression. *Construction Management and Economics, 23,* 295 – 304.

Cheung, F. K.T. & Skitmore, M. (2006). Application of cross validation techniques for modelling construction costs during the very early design stage. *Building and Environment, 41*(12), 1973-1990.

Cheung, F. K. T., Wong, M. W. L., & Skitmore, M. (2008). A study of clients' and estimators' tolerance towards estimating errors. *Construction Management and Economics, 26* (4), 349 – 362.

Council of Tall Building and Urban Habitats (2014). *CTBUH height criteria*. Retrieved from http://www.ctbuh.org/TallBuildings/HeightStatistics/Criteria/tabid/446/language/enUS/Default.aspxhttp://www.ctbuh.org/TallBuildings/HeightStatistics/Criteria/tabid/446/language/en-US/Default.aspx

Dagostino, F. R. & Peterson, S. J. (2011). *Estimating in Building Construction – 7th edition.* Upper Saddle River, New Jersey: Pearson Education.

Fortune, C. & Lees M. (1996). The relative performance of new and traditional cost models in strategic advice and clients, *The Royal Institution of Chartered Surveyors*.

Hwang, S. (2009). Dynamic regression models for prediction of construction costs". *Journal of Construction Engineering and Management, 135*, 360 – 370.

Karshenas, S. (1984). Predesign cost estimating method for multistory buildings. *Journal of Construction Engineering and Management, 110,* 79 – 86.

Kim, G.-H., An, S-H, & Kang, K-I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment, 39*, 1235 – 1242.

Kouskoulas, V. & Koehn, E. (1975). Predesign cost-estimating function for buildings. *ASCE Journal of Construction Division.*

Kwak, Y. H. & Watson, R. J. (2005). Conceptual estimating tool for technology-driven projects: exploring parametric estimating technique. *Technovation, 25*, 1430 – 1436.

Lowe, D. J., Elmsley, M. W., & Harding, A. (2006). Predicting construction cost using multiple regression techniques. *Journal of Construction Engineering and Management, 132* (7), 750 – 758.

Mahamid, I. & Amund, A. (2010). Analysis of cost diverge in road construction projects. *Proceedings of the 2010 Annual Conference of the Canadian Society for Civil Engineering,* 9-12 June 2010, Winnipeg, Manitoba, Canada.

Mahamid, I. (2011). Early cost estimating for road construction projects using multiple regression techniques. *Australasian Journal of Construction Economics and Building, 11* (4), 87 – 101.

McGarrity, R. J. (1988). Parametric estimating: an equation for estimating buildings. Master of Science thesis, Georgia Institute of Technology, Atlanta, Georgia.

Ofori-Boadu, A. N. & Waller, L. (2014). Exploring Conceptual Cost Estimating Models for the World's Tallest Buildings. *Association of Technology Management and Applied Engineering (ATMAE) Conference Proceedings*, *30.* Retrieved from http://c.ymcdn.com/sites/www.atmae.org/resource/resmgr/Event_Materials/2014_Conf.__Proceedings.pdf

Raftery, J. (1984). Models in building economics: a conceptual framework for the assessment of performance. *Proceedings 3rd international symposium on Building Economics, CIB W- 55*, *Ottawa, 3,* 103 – 111.

Road Costs Knowledge System (2002). The World Bank, Transport Unit – TUDTR). Retrieved from http://www.worldbank.org/transport/roads/rd_tools/rocks_main.html

Schexnayder, C. J. & Mayo, R. E. (2003). *Construction Management Fundamentals*. Boston, MA: McGraw-Hill Higher Education.

Skitmore, R. M., & Marston, V. (1999). *Cost modelling*. London: Taylor & Francis.

Sonmez, R. (2008). Parametric range estimating of building costs using regression models and bootstrap. *Journal of Construction Engineering and Management,134* (12), 1011 – 1016.

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

Yu, W.-D. (2006). PIREM: a model for conceptual cost estimation. *Journal of Construction Management and Economics, 24*, 259-270.