An empirical classification of cost overrun in infrastructure projects by using cluster analysis

Fahad Saud A. Allahaim [*fall5762@uni.sydney.edu.au*] and Li Liu

Ph.D. candidate, School of Civil Engineering, Faculty of Engineering and IT, The University of Sydney; College of Architecture and Planning, King Saud University

Senior lecturer, School of Civil Engineering, Faculty of Engineering and IT, The University of Sydney

Abstract

The cost overrun of infrastructure projects potentially poses significant financial risks to the investment stakeholders involved. Based on projects data from across 20 nations in five continents. Studies show that the average cost overrun of infrastructure projects is substantial[1,2]. Yet, over the last several decades, the magnitude of cost overrun of infrastructure projects has failed to improve, suggesting that no significant learning has occurred in the mitigating the extended of cost overrun of infrastructure project.

The aim of the paper is to develop an empirically based classification/taxonomy of cost overrun causes in infrastructure project. We employed a questionnaire survey to elicit the major causes that contribute to cost overrun in infrastructure projects in Saudi Arabia. The frequency and severity of causes that impact on cost overrun are evaluated and ranked. A survey of 160 project mangers of infrastructure projects in Saudi Arabia has been conducted. The data has been analyzed using hierarchical cluster analysis. Results indicated that there is a 4-clusters group. Results indicated that the cluster groups comprised market and regularity uncertainty (30 %), a scope change (10 %), an inadequate planning and control (50%) and site condition (10 %). Each group represents several of causes that lead to cost overrun. Taken together, this cost overrun causes classification/taxonomy provides a structured view that enables an objective evaluation of planning decision methods. Therefore, the model can be used to aid the assessment of cost overrun causes for large infrastructure projects and to effectively mitigate risks of significant overruns.

Introduction

Cost overruns in infrastructure projects are common around the world. High profile examples include: the Wembley Stadium that experienced a 50% cost overrun; and the Scottish Parliament Building that was over three years late and experienced more than 900% cost overrun[3]. In Australia, the Western Australian Perth Arena had an original contract value of AUD 168 million, but a cost overrun of more than three times this amount[3]. According to Flyvbjerg[2,4] the average cost overrun for infrastructure large-scale projects could range from 20.4% to 44.7%; and 90% of projects have cost overruns worldwide, also cost overrun is found across 20 nations and five continents. Over the past 70 years, there have been no systematic improvements in cost overrun of infrastructure projects[5].

Various causes of cost overruns have been identified. Studies have shown that technical factors lead to cost overruns, including lack of experience, project size, mistakes in design, overall price fluctuations, inaccurate estimations[6]. Love and others[3] conducted a study on the causes of cost overruns via case studies on a hospital and a school. They found that technical factors (such as design errors) are the major causes leading to cost overruns.

On the other hand, According to Flyvbjerg et al.,[7]however, there are two basic reasons why projects experience cost overruns. Firstly, optimism bias encapsulates the systematic propensity of decision makers to be over-optimistic about outcomes of planned actions. Secondly, strategic misrepresentations are the misleading actions used in politicisations and economics, and by planners, to ensure projects proceed[8]. Traditional estimation practices have been shown to be particularly vulnerable to these detrimental effects, resulting in poor estimation accuracy in previous studies[9].

It is apparent that there are a large number of causes of overruns and many share similar patterns of impact on overrun costs. Therefore, it will be functionally useful and conceptually meaningful to develop a classification/taxonomy of causes based on their impact on the overruns of infrastructure projects. We identified the frequent causes through reviewing of empirical literature on the cost overrun of infrastructure projects. Based on the empirical study

conducted in Saudi Arabia,a classification/taxonomy of causes has been developed to aid the assessment of cost overrun causes for large infrastructure projects by using cluster analysis. Below, background literature is reviewed and the research method is described. Then, based on the causes identified in the literature review, a classification/taxonomy of causes of cost overrun has been empirically developed. Finally, conclusions are drawn.

Related study

Cost is one of the main considerations throughout a project's lifecycle and can be regarded as a significant parameter of a project and the driving force of project achievement. Despite its proven significance, it is not rare to observe a construction project failing to achieve its objectives within the specified, or even the approximate, estimated cost. Cost overruns vary significantly in scale from project to project. Yet, cost overrun is common to infrastructure projects [10]. Understanding the causes of cost overruns is critical to the success of infrastructure projects. Past studies have found significant, yet common cost overrun of infrastructure projects.

Pickrell[11] carried out a study for the US Department of Transportation covering US rail transit projects with a total value of US$24.5 billion. The total capital cost overrun for eight of the projects was calculated to be 61% ranging from -10 to +106%. Another study by the Auditor General of Sweden[12], covering 15 road and rail projects, revealed that the average cost overrun of eight road projects was 86%. The range for road projects was from -2 to +182%, while the average cost overrun for the seven rail projects was 17%, ranging from -14 to +74%. Another study by Fouracre et al.,[13] carried out for the UK Transport and Road Research Laboratory (TRRL), covered 21 metro systems in developing countries. The outcomes of the study showed that six metro projects had cost overruns above 50%. Two of these projects range up to 500%. Three had cost overruns in the range of up to 100%, and the remaining four ranged up to 50%.

Skamris and Flyvbjerg[14]conducted a study in Denmark, in which they compared the accuracy of cost estimates on large-scale infrastructure projects. The study considered cost estimates of seven tunnels and bridges before the decision was made to build. The major conclusion from this study is that cost overrun of 50–100% is common for larger infrastructures, and that overruns above 100% are not unusual.

Around the globe, many other researchers have been attracted to cost overrun. Asian and African countries have attracted particular attention. In Southeast Asia these researchers are: Kaming et. al.,[15]in Indonesia; Ogunlana et al.,[16] in Thailand, Sambasivan; and in Malaysia, Soon[17]. Chan and Kumaraswamy[18], Chan and Kumaraswamy[19] and Lo et al.,[20] studied cost overrun in Hong Kong, and Acharya et. al.,[21] studied it from a Korean perspective. Chang[22] conducted surveys in the US. In Middle Eastern countries where petroleum and natural gas exports have played an important role in the economy, researchers are: Faridi and El- Sayegh[23] in UAE, Koushkiet al.,[24]in Kuwait.

The studies on causes of overrun have identified a wide spectrum of causes. Frimpong et al.,[25]identified 26 factors that cause cost overruns in the construction of ground water projects. They found that, according to the contractors and consultants, monthly payment difficulties were the most important cost-overrun factor. Owners, however, ranked poor contractor-management as the most important factor. Although there were some differences in viewpoints among the three groups surveyed, there was a high degree of agreement among them with respect to their ranking of the factors. The overall ranking results indicated that the three groups felt the major issues which can cause extreme groundwater project-cost overruns in developing countries are: monthly payment difficulties; poor contractor management; poor technical performances; material procurement; and escalation of material prices.

In Kuwait, Koushki et al.,[24] did a study in which cost increases in the construction project was examined. The study found the three most important causes of cost overruns are contractor elide, material related problems and owners' financial constraints. Other studies have identified four of the most important factors that cause cost overruns as: design changes; inadequate planning; unpredictable weather conditions; and fluctuations in the cost of building materials[26,27].

In Africa, Frimponget al.,[25]conducted studies in Ghana, as did Mansfield et al.,[28] and Aibinu and Odeyinka[29] in Nigeria. In Vietnam, large-scale projects were studied by Long et al.,[30] to identify project success factors, and by Long et al.,[31] to identify ordinary and general issues. Regarding these issues, the Vietnamese government declared the infrastructure project cost-overrun issues as the biggest "headache"[32] in recent times, especially with government-related funded-projects[32].

An Empirical classification of cost overrun in infrastructure projects by using cluster analysis

Skamris et al.,[33] concluded that in most previous studies, technical factors such as changes in design and technological innovation could be explained as causes of cost overruns. However, there remains a considerable portion of divergence that cannot be clarified by technological causes alone[1]. In fact, Wachs[34] pointed out that the probable cause of cost overruns in infrastructure projects is due to the inaccuracy of cost forecasts. Love et al.,[3] did a study of the cause cost overruns within two case studies (hospital and school) in Australia. He found that the technical factors (such as design error) are the major issue that lead to cost overruns.

Flyvbjerg et al.,[7] argues about the main causes of the cost overruns. They postulate that these causes affect projects through the life cycle, and are due to misinformation in policy and the management of the project. Why projects experience cost overruns is firstly due to optimism bias (appraisal optimism) that encapsulates the systematic propensity of decision makers to be overoptimistic about outcomes of planed actions. Secondly, they relate to the strategic misrepresentation (lying) that misleads actions used in politicisations and economics, and by planners to ensure the projects proceed[8]. In addition, Flyvbjerg[9] acknowledge other causes such as project size and location, however, he conclude that optimism bias and strategic misrepresentation are the main causes of cost overrun and thereby he did substantial research, which contributed to improving understanding of the reasons for the infrastructure cost overruns, by collaborating with the optimism bias and strategic misrepresentation.

Kahneman and Tversky[5] developed the theories of reference class forecasting, Flyvbjerg and COWI[35] developed the method for its practical use in policy and planning. They argue that optimism bias and strategic misrepresentation can be measured by a forecasting method "reference class forecasting" (RCF) based on decisions made under uncertainty. By taking the outside view, the RCF approach mitigates optimism bias and strategic representation. Therefore, the RCF technique is by passes human bias by cutting directly to outcomes. It completely ignores the details of the project at hand (e.g. government regulations, the project size, the quality of the contractor management team, plan changes, priority on construction deadlines, etc.) and it involves no attempt at forecasting the event that influences the project's future course.

As it is appeared in the literature of cost overrun causes there are two main schools of thought; technical

school and deception (psychological and political-economic) school. Technical school focus (inside vies) on causes that lead to cost overruns based on how the work is done, these causes are, for example, a lack of experience, the size of the project, mistakes in design, overall price fluctuations, inaccurate estimations, government regulations, the project size, the quality of the contractor management team, plan changes, priority on construction deadlines, completeness and the project information timelines, the experience of the estimators, and bidding conditions; project characteristics, past data on similar types of projects, and the process of estimating.

Deception school developed, which they published various papers on cost overrun causes for infrastructure project and widely received and citied their outcome. The deception school (outside view) causes are optimism bias and strategic misrepresentation. It is clear that the different opinion about the critical factors of the cost overrun is controversial. In addition, the limitation of understanding cost overrun causes creates differences in mitigating the causes effectively. Therefore, it is important to develop a classification of cost overrun causes to reduce the complexity of causes, and to facilitate effective understanding in management of such causes.

Research Design

Data Collection

The survey method for data collection was used through distributing questionnaires and conducting interviews to classify the causes. The survey was conducted in Saudi Arabia. The construction boom in infrastructure projects, which started in 2005, is expected to go through a period of accelerated growth over the next few years, with a value of projects estimated at US $629 billion[36]. On other hand, there is a lack of research on cost overrun in infrastructure projects in Saudi Arabia in the literature. This study fills this research gap.

The questionnaire was directed towards three groups in both public and private organisations; owners, consultants and contractors, who are they involved in infrastructure projects (project mangers). There are three main organisation that dealing with engineers, and contractors and consultants in Saudi Arabia. These organisations are; Saudi Council for Engineers (SCE), Ministry of Municipal and Rural Affairs (MoMRA) and Chamber of Commerce and Industry (CoCl). Their databases were used to distribute the questionnaire

and also to gather some information (contact details) about the participants. The sample selected for each of the three groups is described below as: owners comprising the government agency (key decision-makers) responsible for the projects, consultants working in the infrastructure projects (project managers), contractors who are involved in the infrastructure projects (project managers). Because of the limitation number of project managers who involved of infrastructure projects in Saudi Arabia and registered at Saudi Council for Engineers (SCE), also have experience in infrastructure projects, the sample random number targeted in the study was 500 participants.

The selection of the personnel involved in the interviews was based on their knowledge and work experience in managing infrastructure projects. For example, the Regional Director of the Ministry of Higher Education or, and routinely deals with infrastructure projects and is currently managing the operational contracts of university projects within the region. The personnel involved in the interviews then verify the possibility of the cost overrun causes that identified from the literature review, as occurring in the projects.

Survey

Based on comprehensive review of the relevant literature, main cause of cost overrun in infrastructure projects and their relative reflect are measured. The questionnaire poses specific questions to the respondents' most recently completed infrastructure projects (e.g. education, health, transportation, water, and power) with a contract value over 80 million Saudi Riyals (US $20 million), excluding operation and maintenance cost. The questionnaire consists of three sections: general information about the participant's experience; causes of cost overrun; and the frequency and severity of each of these causes, including the extent of cost overrun, respectively. The questionnaire were distributed online through survey-monkey website in two languages; Arabic and English.

The first section contains questions about participants and their organisation, work experience, academic qualifications, the number of projects constructed within 20 years, location of these projects with type of the projects and the experience with cost overruns through that period. In the second section, the participants were asked to scale the frequency of 41 cost overrun causes using Likert scale response Anchors: (Never (N)=1, Occasionally (OC)=2, Sometime (S)= 3, Often (O), =4, Always (A)= 5).

Furthermore, they were asked to scale the severity of the same causes within the following scale: (No significant (NS)=1, some effect (SE)=2, Moderate (M)=3, Significant (S)=4, Extremely significant (ES)=5). They are also asked about their most recent involvement in a project regarding the overall major causes of cost overrun. The last section of the questionnaire elicits general comment in reference to the study.

In parallel, interviews were conducted with 15 key personnel from the industry in order to get their personal opinions and views about the possibility cost overruns causes in infrastructure projects. The questions were asked to the participants were the same questions in the questionnaire. The interviews were conducted with, 3 project directors, 8 project mangers 4 project mangers assistance.

Descriptive statistics

Questionnaires were distributed to 500, 391 potential responds it is identified from the database. 160 (32%) returned completed questions. 23% of the responds are owners, 52% contractors and 25% consultants. Akintoye and Fitzgerald[37] the respond rate of this study is comparable to 42% of their survey study which about cost estimating practices in the UK, Vidogah and Ndekugri [38] 27% response rate to their survey, improving the management of claims on constriction; and Shash[39] 28.3% responded rate received, which his is study about factors considering in tendering decision by top UK contractors. Moser and Kalton[40] claimed that a survey could be considered biased if the response rate is less than 30%.

The average age of respondents is being 40 years old and over. Most respondents have bachelor degree (58.8%) while the lowest responds have diploma (6.9%). The respondents that have postgrad qualification come second with (34.4%). There is 36 % of responds has experience in infrastructure project over than 10 years.

The average cost overrun of power and health projects is 60%, transport and water projects is 40% and education projects overall cost overrun is 30%. The responsibility of occurring the cost overrun in infrastructures were; 44% owners, 34% contractors, 20% consultants and 2% third party (e.g. other stockholders, changing of government regulation).

The top 10 causes were reported ranked by the computation of importance index are (1) market conditions (materials and labours), (2) design

changes, (3) practice of assigning contract to lowest bidding, (4) delays (in decision making and approval of drawings), (5) design error, (6) deficiencies in the infrastructure, (7) additional works and rework, (8) slow payment of completed works, (9) change in the scope of the project, and (10) changes in material specification and type.

Cluster Analysis

Cluster Analysis (CA) or clustering is a statistical method that consists in assigning a set of several objects into groups (called cluster) so that the objects in the same cluster are more similar to each other than to those in other clusters[41], because the number of groups is usually unknown before the analysis. Clustering involves several distinct steps. The most important steps are the definition of a suitable distance between objects to measure their closeness or similarity, and the choice of a clustering algorithm that we will be applied on the data. In this way, it needs to answer the following questions: how do we define closeness or similarity?, how do we group cost overrun causes?, how do we validate the groups given by the clustering algorithm?, and to finish, how do we visualize and define these groups?. The clustering process consists in four steps: preparing the data, determining the number of clusters, testing the cluster solution hierarchal cluster and validating. R project (version 3.0.2) for statistical computing and graphics is used for the analysis.

The raw data contains 160 observations and 41 variables (causes). These data are 1 to 5 (by ordinal scale), which will be used to measures the closeness in terms of impact on cost overrun. In addition, there are many clustering investigations that have starting point with the data, which is an *nxn* one-mode matrix that reflects a quantitative measure of closeness[41]. Therefore, a data matrix to measure the distance and similarity (proximity) between causes is needed.

Kendall's Tau correlation matrix is used as data collected in this study is ordinal, in R function (cor) was used as correlation plot as showed in figure 1. The sign of the correlation coefficient defined on the range -1 <r +1 measures the relationship. The value '1' reflects a stronger positive relationship and the value '-1' reflects a stronger negative relationship, which is that measures the correlation between two sets of ranking and the equation is written as;

$$Tjk = \frac{Sjk}{(0.5)(P)(P-1)} \qquad (1)$$

Where; *Tjk* = Kendall's Tau correlation coefficient, *Sjk*= the accumulated value of (concordant pairs - discordant pairs), *P* = the number of rows in the data.
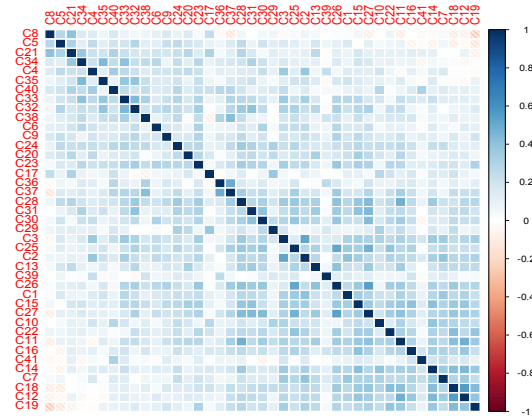

Figure 1: correlation matrix

After the data prepared and the method has been chosen, it requires standardize each variable to reduce bias. This is because to remove arbitrary affect on similarities among objects and to makes attribute contribute more equally to the similarities among each variable. To standardize a variable, for each observation, the mean divided it by the standard deviation. In R, there is a function (scale) that will do standardize as explained in equation 2. This will make the sample have mean 0 and standard deviation 1.These methods suggested by Art and Gnanadesikan[41].

$$z = x - \mu/\sigma \quad (2)$$

Where, z = z score, x = measured value, μ = mean, and σ = standard deviation.

The determination of the number of clustering is based on the data available. Comparing the within sum of squared error (SSE) for each cluster solution (number of clusters) is one of the most common method of choosing the most appropriate cluster solution. The definition of within SSE is the sum of the squared distance between each object of a cluster and its cluster centroid. Therefore, SSE is the main measure of error. Generally, as the number of clusters increases, the within SSE should decrease because clusters are, by definition, more and more small[42]. A plot of the within SSE against a series of sequential cluster levels can provide a useful graphical way to choose an appropriate cluster level. An appropriate cluster solution could be defined as the solution when the reduction of within SSE slows dramatically. This produces an "elbow" in the

plot of within SSE against cluster solutions. In the plot shown below (figure 2), there is an "elbow" at the 4 cluster solution suggesting that solutions over 4 clusters do not have a substantial impact on the total SSE. Therefore, the "elbow" in the plot is extreme and using hierarchical clustering will test that.

$$SSE = \sum_{i=1}^{n}(y - y)^2 \quad (3)$$

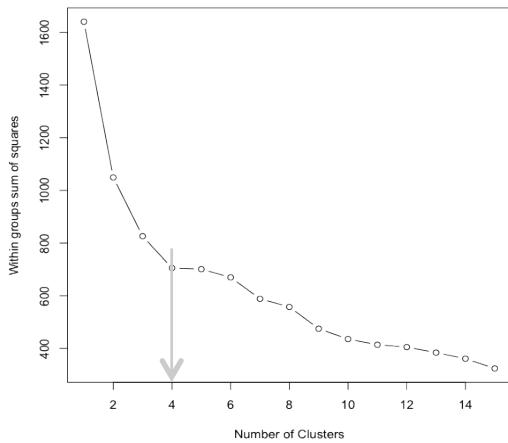Where; SSE = sum of square error, y =observations and y = mean

Figure 2 : Elbow plot for the cluster determination

The next step is testing the cluster solution using hierarchal cluster. The reason that we select hierarchal cluster is because it can test the cluster solution, which is different form K-means clustering that need to specify the cluster solution before start the analysis. This means that hierarchal cluster does not require knowing or specifying the number of cluster, which will specify the number of clustering when it performed.

Hierarchical clustering procedures characterized by a tree-like structure built during the analysis. Most hierarchical techniques fall into a category called agglomerative clustering. In this category, clusters are consecutively formed from objects. Initially, this type of procedure starts with each object representing an individual cluster. These clusters are then sequentially merged according to their similarity[41]. First, the two most dis/similar clusters (i.e., dissimilarity are those with the smallest distance between them similarity are those with the biggest distance between them) are merged to form a new cluster at the bottom of the hierarchy. Examples of distance measure are Euclidean, Manhattan, or Correlation distance. In the next step, another pair of clusters is merged and linked to a higher level of the hierarchy and so on until all the observations are in the same cluster. In this way, we

need to choose a criterion to determine how elements are merged, that depends on the cluster structure. The main linkage criteria used are Single, Complete, Average and Ward's linkage. This allows a hierarchy of clusters to be established from the bottom up. I therefore concentrate on the agglomerative clustering procedures. However, before we discuss these, we need to define how similarities or dissimilarities are measured between pairs of objects.

Select a Measure of dissimilarity or Distance; The most straightforward and generally accepted way of computing distances between objects in a multi-dimensional space is to compute Euclidean distances to generate a dissimilarity matrix, which depends on information value and the nature of the variables describing the objects to be clustered. Hence, in this analysis, we choose the Euclidean distance method[41]. The Euclidean distance is the square root of the sum of the squared differences in the variables' values.

$$D_{\text{Euclidean}}(A, B) = \sqrt{(Xa - Xb)^2 + (Ya - Yb)^2} \quad (4)$$

Select a Clustering Algorithm; after having chosen the distance or similarity measure, we have to decide which clustering algorithm will be applied. Rosmesburg (1984) recommended to use "unweighted pair-group method using arithmetic averages" (UPGMA) which that as same as "average" agglomerative method for ordinal data. This method is very efficient and widely used[41]. The average or UPGMA method considers only distances between pairs of groups in different clusters. The distance of two cluster x,y is defined as the average distance between any one element of x and any one element of y. This method can[41]be used with any kind of dis/similarity or distance measure between groups. This method is implemented in the function hclust(), included with the base distribution of R. Figure 3 shows the hierarchical cluster (tree) that generated from R with a four clean rectangular. Each cluster contains the causes of cost overrun that related to each other in the same cluster.

$$D_{KL} = \frac{1}{n_K n_L}\sum_{i\in C_K}\sum_{i\in C_L} d(x_i, x_j) \quad (5)$$

Where; $d(xi,xj)$ is the distance between objects and ; K and L are two sets of objects (clusters); nk and nL are the numbers of objects in clusters K and L respectively.

An Empirical classification of cost overrun in infrastructure projects by using cluster analysis
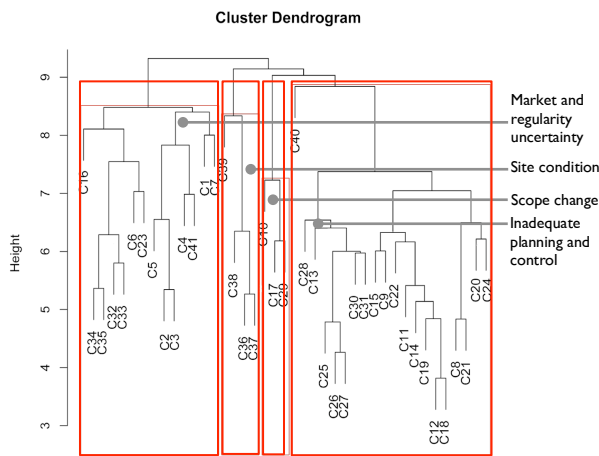

Figure 3: Hierarchical clustering (cut tree)

The next step is validation by using p-value. Cluster analysis is a technique to inspect the similarities/dissimilarities between objects. Hierarchical clustering generates a dendrogram, which contain clusters show the similarities/dissimilarities based on matrix computed by data. It provides detailed information on the relationships between objects. However, it is not clear how strong these clusters are supported by the data. The question is, "How accurate are these clusters?"

To answer the above question, pvclust package (pvclust) is used (available in CRAN packages) for a statistical software R[43]. This package is for assessing the uncertainty in hierarchical cluster analysis. For general statistical problems pvclust can be used easily. Pvclust calculates probability values (p-values) for each cluster using bootstrap resampling techniques. Bootstrap resampling replicates data by resampling from the data itself. It is randomly chosen observations from the original data with replication[44,45]. The procedure is repeated many times. There are two types of p-values, which are approximately unbiased (AU) p-value and bootstrap probability (BP) value. AU p-value is computed by multiscale bootstrap resampling, which is a better approximation to unbiased p-value than BP value computed by normal bootstrap resampling. Therefore, we measure the accuracy of these clusters as p-values, which ranges from 0 to 1.If the p-value of a cluster (or a hypothesis) is less than α, say smaller than 5%, the cluster is rejected at the α level of significance. That indicates how strong the cluster is supported by data[46,47].

In the analysis the number of bootstrap replicates is (nboot) = 10,000 is used to reduce the standard error

(SE). Figure 4 shows the dendogram that generated from the script 5.6,it can be seen that in the figure a four rectangular have AU value of 0.99 or greater, Therefore, a cluster with AU p-value $\geq$ 0.95, the hypothesis is rejected with significance level 0.01 and 0.0, which indicates how strong the cluster is supported by data. Hence, it suggests having four clusters.
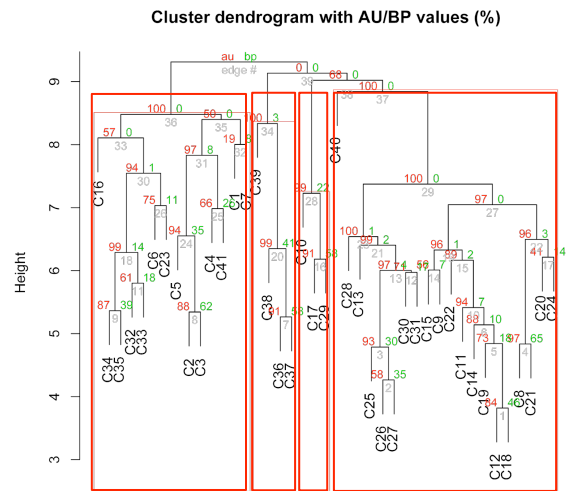

Figure 4: Hierarchical clustering, Values at branches are AU p-values (left), BP values (right), and cluster labels (bottom). Clusters with AU $\geq$ 95 are indicated by the rectangles.

Discussion of result

The determination of cluster number suggested a four clusters solution. The test of cluster level supported four clusters. Each group contains of objects, these objects have relationship between each other. These group are economic uncertainty, uniqueness, pace and complexity.

Inadequate market and regularity

First group represents the causes of Inflation, monthly payment difficulties form agencies (e.g. contractor, owner), cash flow during construction, slow payment of completed works, market conditions (materials and labours), fluctuation in money exchange rate, deficiencies in cost estimates prepared by public agencies, failure to price in some risks, high interest rate charged by bankers on loans, fraudulent practices, obstacles from government, political complexities, laws and regulatory framework, practice of assigning contract to lowest bidding. This group can be defined as the chance or speculation that the cost will change, whether directly or indirectly. Also, it refers to the

possibility that a specific commodity is unstable. That can be seen when the prices of labor, and materials required to construct a project fluctuate unpredictably, or when government regulations changed unpredictably, which leads to increase the cost of construction. Some of these causes ranked in the top ten causes about their importance on the project cost overruns. Many large projects are delayed because of payment delay form owner. In Saudi Arabia the problem of the payment delay is not because of insufficient fund but because of the procedure of payment by owner take long time for approval. Owners should have clear plan of payment for each project to pay according to the agreement. Also the owner should consider the assign of low biding carefully as must of this is major impact on cost overrun.

Site Condition

Second cluster group represents causes weather conditions, site constraint, problem with neighbours, heritage material discovering. These causes are related to environment issues, social and culture impact. This means that these issues can lead to increases the uncertainty of tasks and outcome, which can make planning and estimating difficult. The increasing of environmental requirements is impact significantly on construction operations. For example, the environmental issues (i.g.unexpected geological conditions) can lead to increases the uncertainty of tasks and outcome, which can make planning and estimating difficult. The measurements of the environmental issues have had many consequences for the construction industry. The Economic Commission for Europe[48] believed that the environmental issues are resulted on an increase in costs and time of the process of design, planning and construction life cycle. The public exposures of the project (i.g. protests) over their environmental impacts have had adverse impact on the project itself, which led to the abandonment or cancellation of numerous projects, which were at the stages of completion[49]. In the UK for example, the alliances of residents of districts traversed by the proposed route and non-governmental organizations add a pressure to modify the design of British Rail's high-speed rail link to the Channel Tunnel (to put 36.8 of the 108.8 km in tunnels, with just over 35 km in cutting), it is estimated to have added some US$1.4 billion to the total construction cost[49].

Scope change

Third cluster group represents causes of design changes, additional works and rework, change in the scope of the project. These causes ranking as well in

top ten according to the importance. In Saudi Arabia, these days the construction industry going faster swelling, and because of that the scope and design stage of the project did not consider carefully, which leads to neuromas of design issues through the constriction. Inspection and design approval stage considered poorly. Unclear project scope from owners lead to unrealistic designs, which that results in projects delayed and therefore cost overrun. Very clear scope defined lead to generate an accurate design, which that lead to less cost and time[19].

Planning and control uncertainty

Fourth cluster group represents causes of deficiencies in the infrastructure, labours, Insurance, work security or workers' health problems, lack of experience of project (e.g. location, type), contractor's poor site management and supervision, shortage of site workers, unrealistic contract duration and requirements imposed, strategic misrepresentation, incorrect planning and scheduling by contractors, Late delivery of materials and equipment, changes in material specification and type, design error, project size, Inadequate specifications, waste on site, equipment availability and failure, poor financial control on site, lack of constructability, inadequate modern equipment (Technology), optimism bias, delays (e.g. in decision making, in approval of drawing. All of the causes related to project planning and control, which composed of most critical causes in large projects in Saudi Arabia. Contractor's poor site management and supervision is big issue in Saudi Arabia and represented weakness. Training skilled, updating knowledge for workers are really important for large project. Contractor and consulate selection stage must consider carefully[50]. Coordination, communication and control of management are really important for construction projects. Setting communication channels between parties that run effectively to deal with difficulties arising during implementation with advance information technology will reduce the misunderstanding and increase the quality of work[50].

Comparing the finding with other study

Long et. al.[50] did study that categorize the causes of cost and time overrun for constriction industry in Vietnam using factor analysis technique, which they identified 7 groups: slowness and Lack of constraint, incompetence, design, market and estimate, financial capability, government, and worker.

Abdul Rahman et al[51] carried out study about modeling causes of cost overrun in large construction projects

with partial least square-SEM approach. The study categorize the cost overrun in 7 groups based on the literature and then test these groups amongst contractors involving in large construction projects in Malaysia. These groups that investigated are contractor's Site Management related factors, design and documentation related factors, financial Management, information and communication, Human Resource, Non- human; and project management and Contract Administration. They found Based on hierarchal model for assessing causative factors and cost overrun that contractor's site management related factor strong effect on cost overrun.

Shenhar and Dvir[52], proposed an approach called "adaptive project management", which has two major aspects; project success criteria and diamond model (Novelty, Technology, Complexity, Pace model) by using a typology approach, which can help identify the project type and style. For example, a project that is "derivative", "medium-tech", "system" and "fast/competitive" needs to be managed differently than a project that is "breakthrough", "high-tech", "system" and "blitz". The framework is for planning and managing of technology or product-line projects such as luggage's systems in airport, air conditioning system in building and car projects[52].

Comparing with our study; the slowness and Lack of constraint, incompetence groups of Long et al[50] study reduces into two main groups with our finding, which are inadequate planning and control and scope change groups. Also that supported by Abdul Rahman et al.[51]. Market and Estimate, financial capability, government, and worker of Long et. al.[50] finding is reduced into one group of our finding that is market and regularity uncertainty. This study add new group that been not identified in the literature which are site condition. Also based on cluster analysis our study reduces the groups to 4 groups with a comprehensive definition that can apply to any project in different location. Therefore, The model delivers a conceptual overview that can be used by project managers.

Implications for practice

We use inadequate planning and control, scope change, site condition and market and regularity uncertainty definitions to measure the cost overrun of assertions, components, and the rationale as a whole. Taken together, this cost overrun causes model provides a structured view that enables an objective evaluation of planning decision methods. Therefore, the model will reduce the complexity of understanding the causes that lead to understand large number of causes of overruns

may share similar patterns of how it impact on overrun, and that help to facilitate effective management of such causes.

Conclusion

The purpose of this paper was to develop an empirical classification to aid the assessment of cost overrun causes for large infrastructure projects, to identify the major types and to measure the relationship between causes and overrun. Since there are many studies identifying various causes, we synthesized the empirical literature on infrastructure project cost overrun causes and analyzed the frequency of cost overrun causes. Based on developing an empirical classification via a survey that has been conducted in Saudi Arabia. We have used the survey data for the analysis. The cluster analysis is composed of four phases: preparing the data, determine the number of clusters; test the cluster solution by suing hierarchal cluster and validation by using p-value. Data were first prepared by inspected the correlation to generate a data matrix. Then, determine the number of cluster based on SSE. Then, using hierarchal clustering to test the cluster solution. Finally a validation using P-value that developed and add-on R package by Shimodaira and Hasegawa[43].

The potential contribution of this study is in identifying an empirically derived classification of cost overrun causes. The determination of cluster number suggested a four clusters solution. The test of cluster level supported the four clusters method hierarchal clustering. Each group contains of objects, these objects have relationship between each other. Hence, the cluster represents dimensions of cost overrun. Finally, the P-value shows the cluster is strong with four cluster solutions with 0.99 of greater, which that respects the value of 0.01 or low. These cluster groups are "inadequate planning and control", "scope change", "site condition" and "market and regularity uncertainty". This finding helps to identify the main reason why there is cost overrun. Also, it helps to mitigate that problem in right way and can be included on a cost estimation method. Therefore, it can be used in the early stage of the project. In the future work will develop the groups and evaluate them based on the literature and use the 4 clusters for cost forecasting.

Notes

[1]Odeck, J. (2004) 'Cost overruns in road construction what are their sizes and determinants'. *Transport Policy*, 11(1), pp. 43–53

An Empirical classification of cost overrun in infrastructure projects by using cluster analysis

[2]Flyvbjerg, B. (2009), 'Survival of the unfit test: why the worst infrastructure gets built—and what we can do about it'. Oxford Review of Economic Policy, Volume 25 (3), pp.344–367.

[3] Love, P., Edwards, D., &Irani, Z., (2011), 'Moving Beyond Optimism Bias and Strategic Misrepresentation; An Explanation for Social Infrastructure Project Coats Overruns', *IEEE Transaction on Engineering Management.*

[4] Flyvbjerg, B., Holm, M.S. and Buhl, S. (2005) 'How (in)accurate are demand forecasts in public works projects?'.*Journal of the American Planning Association*, 71(3), pp.131–146.

[5]Bruzelius, N., Flyvbjerg, B. and Rothengatter, W. (2002) 'Big decisions, big risks, improving accountability in megaprojects'. *Transport Policy*, 9(2), pp.143–154.

[6]Memon, A., Abdul Rahman, I., & Abdul Aziz, A., (2011), 'Preliminary Study on Causative Factors Leading to construction Cost Overrun'.*International journal of Sustainable Construction Engineering and Technology*, 2 (1), pp. 57-71.

[7]Flyvbjerg, B., Holm, M.K. and Buhl, S.L. (2004) What causes cost overrun in transport infrastructure projects? *Transport Reviews*, 24(1), pp.3–18.

[8] Flyvbjerg, B. (2006),'From Nobel Prize to project management: getting risks right'. *Project Management Journal*, 37(3), pp. 5–15.

[9]Flyvbjerg B., Holm M.S., & Buhl S., (2002), 'Underestimating costs in public works projects', *APA Journal*, Summer, Vol.68. No.3,pp. 279-295.

[10]Azhar, N., Farooqui, R. U. and Ahmed, S. M. (2008). 'Cost overrun factors in construction industry in Pakistan'. *Proceeding of first international conference on construction in developing countries* (ICCIDE-1), Karachi, Pakistan, 4-5 August. Pp. 499-508.

[11]Pickrell, D.H., (1990). 'Urban rail transit projects: forecasts versus actual ridership and cost'. US Department of Transportation, Washington DC.

[12] Auditor General of Sweden. (1994) *Riksrevisionsverket.*[The Swedish National Audit Bureau.] *Riksrevisionsverket.*, Vol. 23. Auditor General of Sweden, Stockholm,Sweden.

[13]Fouracre, P.R., Allport, R.J.,& Thomson, J.M., (1990). 'The performance and impact of rail mass transit in developing countries'. TRRL Research Report 278. Transport Research Laboratory, Crowthorn, Berkshire, England.

[14]Skamris, M.K., &Flyvbjerg, B., (1996). 'Accuracy of traffic forecasts and cost estimates on large transportation projects'.Transportation Research Record, pp.15-18.

[15]Kaming, P.F. ,Olomolaiye, P.O. , Holt, G. D. , & Harris, F. C. (1997). 'Factors Influencing Construction Time and Cost Overruns on High-rise Projects in Indonesia'.*Construction Management and Economics*, 15(1), pp. 83-94.

[16]Ogunlana A.O., &Promkuntong K., (1996),' Construction delays in a fast-growing economy: comparing Thailand with other economies', *International Journal of Project Management*, 14(1), pp.37- 45.

[17]Sambasivan, M. and Soon, Y.W. (2007), 'Causes and effects of delays in Malaysian construction industry', *Int. J. Project Management*,Vol. 25, pp. 517-526.

[18]Chan D.W.M. &Kumaraswamy M.M., (1995), 'A study of the factors affecting construction durations in Hong Kong', *Journal of Construction and Economics*, 13, pp.319-333.

[19]Chan D.W.M. &Kumaraswamy M.M., (1997), 'A comparative study of causes of time overruns in Hong Kong construction projects', *International Journal of Project Management*, 15(1), pp. 55-63.

[20] Lo, T.Y., Fung, T.W.H. & Tung, K.C.F. (2006), 'Construction delays in Hong Kong civil engineering project', *J. Constr. Eng. Management*, ASCE, 132 (6), pp. 636-649.

[21]Acharya, N.K., Lee, Y.D., and Im, H.M. (2006). "Investigating delay factors in construction industry: A Korean perspective." *KoreanJournal of Construction Engineering and Management*, Vol. 10, pp. 177-190.

[22] Chang A.Shing-Tao (2002). 'Reasons for cost and schedule increase for engineering design projects' *J. Management Eng.,* ASCE, (18)1, pp. 29-36.

[23]Faridi, A.S. and El-Sayegh, S.M. (2006). 'Significant factors causing delay in the UAE construction industry', *Construction Management and Economics*, (24)11, pp. 1167-1176.

[24]Koushki, P. A., Al-Rashid, K &Kartam, N. (2005). 'Delays and cost increases in the construction of private residential projects in Kuwait'. Construction Management and Economics, 23 (3), pp 285-294

[25]Frimpong, Y., Oluwoye, J., and Crawford, L. (2003). 'Causes of delay and cost overruns in construction of groundwater projects in a developing countries; Ghana as a case study.'*Int. J. ProjectManagement*, Vol. 21, pp. 321-326.

[26]Kaming, P.F. ,Olomolaiye, P.O. , Holt, G. D. , & Harris, F. C. (1997). 'Factors Influencing Construction Time and Cost Overruns on High-rise Projects in Indonesia'.*Construction Management and Economics*, 15(1), pp83-94.

[27]Chimwaso D.K., (2000), 'An evaluation of cost performance of public projects: Case of Botswana', *Proceedings of the 2nd International Conference of the CIB*, http://buildnet.csir.co.za/cdcproc/docs/2nd/chimwaso_dk.pdf.

[28] Mansfield, N.R., Ugwu, O.O., and Doran, T. (1994), 'Causes of delay and cost overruns in Nigeria construction projects', *Int. J. Project Management*, (12) 4, pp. 254-260.

[29]Aibinu, A.A. and Odeyinka, A. (2006). "Construction delays and their causative factors in Nigeria." *J. Constr. Eng. Management,* ASCE, (132) 7, pp. 667-677.

[30] Long, N.D., Ogunlana, S.O., and Lan, D.T.X. (2004).'A study on project success factors in large construction projects in Vietnam', *Journal of Engineering, Construction and Architectural*, (11) 6, pp. 404-413.

[31]Long, N.D., Ogunlana, S., Quang, T., and Lam, K.C. (2004). 'Large construction projects in developing countries: a case study Vietnam', *Int. J. Project Management*, Vol. 22, pp. 553-561.

An Empirical classification of cost overrun in infrastructure projects by using cluster analysis

[32] Le-Hoai, L., Lee, Y., & Lee, J.,(2008). 'Delay and Cost Overruns in Vietnam Large Construction Projects: A Comparison with Other Selected Countries', *KSCE Journal of Civil Engineering*, 12(6), pp. 367-377

[33] Skamris, M.K., &Flyvbjerg, B., (1996). 'Accuracy of traffic forecasts and cost estimates on large transportation projects'.Transportation Research Record, pp.15-18.

[34] Wachs, M., (1990), 'Ethics and advocacy in forecasting for public policy', *Business and Professional Ethics Journal* 9 (1–2).

[35] Flyvbjerg, B. and COWI. (2004) *'Procedures for Dealing with Optimism Bias in Transport Planning'*.British Department of Transport, London.

[36] Al-alarbia (2012). "Saudi Arabia eying boom in construction valued at $629 billion" link visited on November, 30th 2012; < http://english.alarabiya.net/articles/2012/07/28/228936.html>

[37] Akintoye, A. & Fitzgerald, E., (2000), 'A survey of current cost estimating practices in the UK', *Constriction Management and Economics,* 18, pp. 161-172.

[38] Vidogh, W. and Ndekugre I. (1998) Improving the management of claims on construction contracts: consultant's perspective, *Construction Management and Economics*, 16, pp.363 – 372.

[39] Shash, A.A. (1993) Factors considered in tendering decisions by top UK contractors.*Construction Management and Economics,* 11, pp.111-8.

[40] Vidogh, W. and Ndekugre I. (1998) Improving the management of claims on construction contracts: consultant's perspective, *Construction Management and Economics*, 16, pp.363 – 372.

[41] Everitt, B. S., Landau, S., Leese, M., and Stahl, D.,(2011). Cluster Analysis. London, Edward Arnold.

[42] Peeples, M., (2011) "R Script for K-Means Cluster Analysis". Society for American Archaeology style. [electronicdocument]. visited on March 27, 2013 : <http://www.mattpeeples.net/kmeans.html>

[43] Shimodaira,H. and Hasegawa,M. (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics, 17, pp.1246–1247.

[44] Efron,B. (1979) Bootstrap methods: another look at the jackknife. Ann. Stat., 7, pp. 1–26.

[45] Felsenstein,J. (1985) Confidence limits on phylogenies: an approach using the boot- strap. Evolution, 39, pp. 783–791.

[46] Suzuki,R. and Shimodaira, H. (2004), "An application of multiscale bootstrap resampling to hierarchical clustering of microarray data: how accurate are these clusters?" In proceedings by the Fifteenth International Conference on Genome Informatics (GIW 2004), pp. P034.

[47] Suzuki, R., Shimodaira, H., (2006) , "Pvclust: an R package for assessing the uncertainty in hierarchical clustering", Bioinformatics Applications Note journal, Oxford University Press, Vol. 22 No. 12, pp. 1540-1542

[48] Economic Commission for Europe (1986) *Long-term Perspectives for Human Settlements Development in the ECE Region.* (A report prepared by S. Staynov and K. Baumgarter.) United Nations, New Y ork.

[49] Ofori, G., (1992) The environment: the fourth construction project objective?, Construction Management and Economics, 10:5, 369-395

[50] Le-Hoai, L., Lee, Y., and Lee, Y., (2008),"Delay and cost overruns in Vietnam large construction projects: A comparison with other selected countries ". *KSCE Journal of Civil Engineering*, 12, (6), P 367-377

[51] Abdul Rahman, I., Memon, A., Abdul Azis, A., and Abdullah, A., (2013),"Modeling Causes of Cost Overrun in Large Construction Projects with Partial Least Square-SEM Approach: Contractor's Perspective"*Research Journal of Applied Sciences, Engineering and Technology,* 5(6): 1963-1972

[52] Shenhar, A.J. and Dvir, D., (2007) *"Reinventing project management; the diamond approach to succsful growth and innovation"*, Harvard Business School Press, Boston, Massachusetts, USA.