MANCHESTER 1824
The University of Manchester
Manchester Business School

This is the author's version of a work that was submitted/accepted for publication in the following source:

Lowe, D J, Emsley, M W and Harding, A (2006) Relationships between total construction cost and project strategic, site related and building definition variables. *Journal of Financial Management of Property and Construction*, **11**(3), 165-180, eScholarID:1g121 | DOI:10.1108/13664380680001087

This file was downloaded from: https://www.escholar.manchester.ac.uk

© **Copyright 2006 Emerald Group Publishing Limited**

Reproduced in accordance with the copyright policy of the publisher.
Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source.*

Relationships between total construction cost and project strategic, site related and building definition variables

David J Lowe, Senior Lecturer, Centre for Research in the Management of Projects, Manchester Business School, University of Manchester, Manchester, M15 6PB, UK

Margaret W Emsley, Lecturer, Management of Projects Research Group, School of Mechanical, Aerospace and Civil Engineering, The University of Manchester, UK

Anthony Harding, Halcrow Group Ltd, Hammersmith, London

Correspondence Address:

Dr David J Lowe
Centre for Research in the Management of Projects
Manchester Business School
University of Manchester
Manchester
M15 6PB

Tel: 0161 306 4643
Fax: 0161 306 4646
E-mail: david.lowe@manchester.ac.uk

Lowe, D J, Emsley, M W and Harding, A (2006) Relationships between total construction cost and project strategic, site related and building definition variables. *Journal of Financial Management of Property and Construction*, **11**(3), 165-180, eScholarID:1g121 | DOI:10.1108/13664380680001087

Relationships between total construction cost and project strategic, site related and building definition variables

Abstract

There is a paucity of recent literature on the influence of project strategic, site related and design related variables on the cost of construction. This paper seeks to redress this omission by presenting the results of an investigation into the influence of 41 independent variables on both construction cost and client cost. Data were collected from 286 construction projects in the United Kingdom and correlation and test for differences were used to determine the relationships that exist between the dependent and independent variables. The analysis both confirms the strong relationship between construction cost and client cost and between those two measures of cost and GIFA, and establishes other relationships which exist within the data, confirming many of the relationships that had been anticipated from the literature. It also established the ordinal sequence of several nominal variables. These data, therefore, can be confidently used to develop models of the total cost of construction.

Keywords:

Construction cost, client costs, project strategic variables, site related variables, building definition variables

INTRODUCTION

Various models have been developed to estimate (forecast) the cost of buildings (Newton, 1991). These include the application of regression analysis (c.f. McCaffer, 1975; McCaffer *et al.*, 1984; Skitmore and Patchell, 1990; Trost and Oberlender, 2003; and Lowe *et al.*, 2006a),

Lowe, D J, Emsley, M W and Harding, A (2006) Relationships between total construction cost and project strategic, site related and building definition variables. *Journal of Financial Management of Property and Construction*, **11**(3), 165-180, eScholarID:1g121 | DOI:10.1108/13664380680001087

knowledge based expert systems (Brandon et al. 1988 and Brandon, 1990), artificial neural networks (c.f. Elhag and Boussabaine, 2001 and 2002; Emsley et al., 2002, Günaydm and Goğan, 2004; and Kim et al. 2004b), while Kim et al. (2004a) compared cost estimating models based on regression analysis, neural networks and case-based reasoning. Predominately, these models rely on the use of historic (but recent) cost data. In the United Kingdom, the Building Cost Information Service (BCIS) provides details of completed construction projects and their associated costs. While cost advisors may use these data to advise on the cost of a building based on the cost of a similar comparable project, adjusted to reflect any differences, this does not enable any general principles about the relationships which exist between the cost and significant predictor variables to be established.

The aim of this paper is to examine data collected from 286 construction projects in the United Kingdom and to investigate the relationships which exist between the dependent (construction cost and client cost) and independent variables (41 project descriptors) for which data were collated. Ultimately the data were used to develop both regression and neural network models (Lowe *et al*., 2006a and Emsley *et al*., 2002 respectively), however, it is vital to have an understanding of the characteristics of the data to enable the optimum development of such models, while an understanding of the relationships which exist between variables will also enhance the cost advice that advisors can offer their clients.

This, the first of two papers, examines the relationships associated with the two dependent output variables (construction cost and client cost) and project strategic, site related and building definition independent input variables. A second paper (Lowe *et al.* 2006b) discusses the design related variables, with the exception of building definition.

BACKGROUND

According to Morton and Jagger (1995), significant characteristics of buildings include function, size, type of frame, number of storeys etc.; alternatively, Ashworth (2004) suggest the following characteristics: site conditions, building size, wall-to-floor ratio, planning efficiency (ratio of useable space), plan shape, height, storey height, grouping of buildings, buildability and constructional detail. A further source of characteristics is the RICS Building Cost Information Service Standard Form of Cost Analysis, introduced in 1969, which provides elemental cost breakdowns of building costs; elements are defined as "that part of a building which performs a specific function independently of quantity or quality". These include substructure; superstructure, which includes, frame, upper floors, roof, stairs, external walls, windows and external doors, internal walls and partitions, and internal doors; internal finishes: wall finishes, floor finishes and ceiling finishes; fittings and furnishing; services and external works.

Standard texts on building economics and cost planning (for example, Ashworth, 2004; Ferry *et al*., 1999; Jagger *et al*., 2002; Morton and Jagger, 1995; and Seeley, 1995) discuss the implications of various design cost parameters on construction costs. These include: the size and height of the building; plan shape, roofing solutions, for example, its structure, covering and size. Ferry *et al.* (1999) discuss the various measures of cost efficiency of a building shape: wall/floor ratio, plan compactness (POP) ratio, mass compactness (VOLM) ratio, length breadth index and plan/shape index. However, these texts do not provide a measure of the relationship between the various design variables and construction cost or their relative importance, except in a superficial way. For example, Morton and Jagger (1995) state that function and size (floor area) are the 'two most obvious' determinants of construction costs.

Empirical studies into the relationship between building characteristics and construction cost generally tend to be dated. These include that of the Wilderness Group (1964), which investigated the relationship of the design variables: storey height, floor loading, column spacing, and the number of storeys on building cost. Braby (1975) utilised regression analysis to investigate the costs of high-rise buildings. Likewise, Flanagan and Norman (1978) examined the relationship between height and construction price, while Pegg (1984) investigated the effect of location, building function, building height, form of contract and site conditions on tender levels. More recently, Picken and Ilozor (2003) investigated the relationship between height and construction costs of buildings in Hong Kong; Swaffield and Pasquire (1999) analysed the relationship between building form and function, and the cost of mechanical and electrical services; Wing (1999) considered the issue of plan shape complexity and Tan (1999) assessed the influence of height on construction costs.

Further evidence of the influence of building characteristics on construction costs can be gained from examination of the literature on construction cost modelling. For example, Skitmore and Patchell (1990) present a regression model that comprises the following independent variables: gross floor area, number of storeys, air conditioning, contract period and number of tenders. Likewise, the ELSIE system required input information that included the type of client, sector, building status and function, number of staff, GIFA, number of storeys, maximum storeys, basement, wall appearance and detailing, level of fit out, column spacing/open plan area, air conditioning, region, locality, site conditions, ground conditions and construction period Brandon *et al.* (1988).

METHODOLOGY

The initial impetus for the research arose from the paucity of data available that can provide reliable information about the relative costs of using different procurement routes. However, it soon became apparent that this variable cannot be isolated from the other cost significant variables that influence the cost of a building. (Harding *et al*., 1999).

Input and output variables

The data collected were divided into independent input variables and dependent output variables (Emsley *et al.,* 2002).

Two independent variables were identified – construction cost and client costs, which may be combined to give the total cost to the client. A criticism of previous cost models is that they used only the tender price to evaluate cost, whereas, in reality, the cost to the client of a building contract is the final contract sum. This is very rarely the same as the tender price and Corbett and Rowley (1999) suggested that the final account sum should be made available to cost planners, whereas the BCIS, for example, only provide the tender price. The analysis described here has utilized final account figures as the construction cost output variable. In addition, the whole cost to the client includes not only the final contract sum but also the client's additional costs (client costs). These include: planning, building regulation, site investigation, architect, structural/civil engineer, quantity surveyor, planning supervisor, employer's agent, mechanical engineer, and landscape architect fees, plus the client's administrative charge and the cost of resources provided by the client.

An extensive literature review identified numerous predictor (input) variables. These were finally reduced to 41 variables, which it was believed would be known at the early estimating

stage (the stage at which the models are intended to be used). The predictor variables were further categorised as project strategic, site related or design related, and are tabulated in Table 1.

<<<< Insert Table 1 about here >>>>

The variables of time and geographic location were accommodated through the application of the BCIS cost indices to bring all projects to a common location and base date. Also, where projects included external works, demolition, fittings or specialist services, their associated costs were removed from the final account figure, and appropriate proportions were removed from the contract preliminaries and clients' costs (Harding *et al*., 2000). This was done because these costs are subject to wide variation, largely independent of the main variables defining the building. For example, for the projects where data were collected, the cost of external works varied from 1% to 30% of the total contract sum. Such variation makes these features impossible to model accurately and they are more reliably estimated independently.

Data collection

In total, the data collection programme resulted in the collection of 288 full data-sets from predominantly primary sources, supplemented by some secondary data.

Primary sources

The professional collaborators provided a great deal of the data required and contact was also established with organizations, primarily Quantity Surveying and Project Management practices, that were willing to provide data. A data pro-forma was developed to assist both the researchers and, more importantly, those collaborators willing to carry out the data retrieval themselves.

This method of collection provided the great majority of building cost analyses. A total of 39 offices were visited from 20 different organisations across the United Kingdom.

Secondary sources

The BCIS publishes cost analyses for construction projects, which fulfilled the data requirements, except for the following information:

- final account
- actual duration
- quality of building
- clients' external costs
- clients' internal costs

In order to obtain this additional information, a questionnaire was administered and sent to BCIS subscribers, yielding 29 sets of data.

In addition to these questionnaires, data were obtained from a much more extensive mail-shot administered to 1239 practising quantity surveyors, all of whom had been canvassed by telephone, but this yielded only 6 additional projects. A more detailed account of the data collection issues can be found in Harding et al. (2000).

Data representation

The way the predictor variables are represented fell into four distinct groups.

The first of these groups comprised variables which are real numbers, for example, *duration* and *no. lifts.* Where the range of these variables differed by more than one order of magnitude it was more appropriate to use the logarithm of that value, to ensure that the range of values was more evenly distributed.

The remaining variables are categorical variables that represent one of a choice of categories. As a general rule it is best that a single input is used for a variable only when that variable has some meaning as a single variable (Tarassenko, 1998). That is to say, if the value of the variable increases then it must represent an increase of some factor which influences the outcome of the model. For some variables, obtaining such an order was simple. For example, with *site access* there is clearly an order between *unrestricted*, *restricted* and *highly restricted*, inasmuch as an increase in the restriction to access will be expected to cause an increase in cost. Therefore this variable can be represented by a single input.

There were a great many more variables for which no such order was immediately apparent, for example, *internal wall finishes*, where the variable represents the cost of different material combinations which will make up the finish. The value of the input was set to be the standard cost per m^2 of each finish, which provides an order proportional to how much each finish is expected to impinge upon the final building cost.

Finally, there are those variables where it is not possible to ascertain a standard cost and where there is a lack of consensus on comparative costs, so a consistent ordering which would apply in all circumstances cannot be obviously obtained. An example is frame type, which is categorised as '*in situ*', '*masonry*', '*precast*', '*steel*' or '*timber*'. This requires the use of binary input coding (yes/no) which is applied to each possible choice, thus treating each

categorical variable as a series of binary variables, with the disadvantage of increasing the number of variables in the model and subsequently increasing the requirements for data. However, by analysing the data obtained for these variables (as described in this paper), it is possible that a consistent order may be obtained which will considerably facilitate any subsequent data modelling. Again, a more exhaustive description of the data representation issues can be found in Harding et al. (2000).

Analysis

Data analyses were undertaken using the Statistical Package for the Social Sciences (SPSS for Windows, release 12.0.0.). Log transformations were performed for all continuous scale variables to create or approximate a normal distribution in order to satisfy parametric test assumptions. Descriptive statistics and tests for normality were calculated for each dependent variable. The dependent variables were then correlated, using Pearson's product moment and Spearman's rank correlation, with the independent variables. Additionally, partial correlation coefficients were produced controlling for GIFA. Likewise, descriptive statistics and tests for normality were calculated for each independent variable. The continuous scale independent variables were correlated, using Pearson's product moment and Spearman's rank correlation, with the dependent variables. The categorical variables were analysed for differences between subgroups by means of one-way analysis of variance (ANOVA) and its equivalent non-parametric test, the Kruskal-Wallis H test, or by means of the t-test and the Mann-Whitney U/Wilcoxon Rank Sum W test. Relationships between the independent variables were investigated using Spearman's rank correlation. This analysis was exploratory in nature and is, therefore, interpreted by reference to significance levels (two-tailed).

DATA EXPLORATION

Dependent variables

Construction Cost

Four measures of cost were observed: the raw cost (COST), the natural logarithm of the cost (*LN*COST), cost/m^2(COSTM^2) and the natural logarithm of the cost/m^2 (*LN*COSTM^2) As construction cost is to be used as an output for the modelling and will be minimised by least squares, it is important that the cost measure is as close to normal as possible. Descriptive statistics for measures of cost are presented in Table 2, showing the relative values of skewness and kurtosis.

<<<< Insert Table 2 about here >>>>

The high skewness and kurtosis are caused by the fact that most of the projects are of a low cost, while only a few have a very high cost. Thus most of the projects have a cost which is less than the mean (in fact, only 22% are higher than the mean). The skewness is still observed with the cost/m^2, but much less so. Moving to the logs, it can be seen that the selection of a log value decreases the skewness and kurtosis, and subsequently means the data are more closely aligned to a normal distribution (based on Kolmogorov-Smirnov and Shapiro-Wilk tests of normality), suggesting that the log of cost/m^2 would be the most appropriate output for modelling purposes (although it was found that the cost/m^2 model gave the best neural network model (Emsley *et al*. 2002)).

The main correlations (those higher than 0.5) on both the cost and the cost/m^2 are shown in Table 3 and Table 4 which highlight that there is much disparity between those variables which correlate highly with cost and those which correlate highly with cost/m^2. Interestingly,

Elhag and Boussabaine (2001) established similar correlations between tender price and GIFA (0.940), duration (0.600) and number of storeys (0.400).

<<<< Insert Tables 3 & 4 about here >>>>

Given that the correlation between cost and GIFA is so high, it was necessary to try to determine the effect of any correlations between GIFA and the other significant variables on the correlation between those variables and the cost. Therefore, a partial correlation on cost, controlling for GIFA, was performed (Table 5) which shows that when GIFA is used as a control for the correlations, many of the high correlations on cost become much reduced. Only two variables are relatively unaffected: client cost and lifts. It should be noted that only correlations above 0.5 are shown and they are all highly significant at the 0.001 level.

<<<< Insert Table 5 about here >>>>

Client Cost

Descriptive statistics for client cost are shown in Table 6. Note that there are only 67 sets of these data. Client cost correlates well with both cost and GIFA. This suggests that cost/m^2 should be used rather than the raw cost figure, as for cost. The correlation between building size and client costs explains the high values of both skewness and kurtosis. The values of skewness and kurtosis for client cost/m^2 are acceptable (based on Kolmogorov-Smirnov and Shapiro-Wilk tests of normality).

<<<< Insert Table 6 about here >>>>

The main correlations (those higher than 0.5) on both the client cost and the client cost/m^2 are shown in Table 7 and Table 8 which show that the only variables which correlate well with client cost/m^2 are other cost related variables; in addition, many of the highly correlated variables are similar to those appertaining to cost. This is not surprising, as Table 9 shows that cost and client cost are well correlated.

<<<< Insert Tables 7, 8 & 9 about here >>>>

Independent project strategic variables

Descriptive statistics for project strategic related variables are shown in Table 10, with the exception of duration, which is a variable measured on a continuous scale. The number of data sets collected are shown for each category of each variable in respect of both cost/m^2 and client cost/m^2.

To remove the influence of building function on the construction cost/m^2 dependent variable a further variable was generated (Cost/m^2)/Function)); that is, the Cost/m^2 of each project divided by the median Cost/m^2 for its broad function subgroup from the BCIS database to create a proportional value.

<<<< Insert Table 10 about here >>>>

Table 11 gives details of the statistical tests which have been carried out between sub-groups of project strategic variables

<<<< Insert Table 11 about here >>>>

Contract form

As shown in Table 10, contract form is grouped into 6 categories and significant differences between the categories were observed in all cost variables. However, this was also observed with many other cost significant variables, so it is entirely possible that the apparent relationship with cost arises from correlations with other variables.

Duration

Duration is well distributed, having a mean of 42.95 weeks, a range of 9-140 weeks, with skewness and kurtosis of 1.354 and 2.597 respectively. However, the log of duration follows a distribution that is very close to normal, having skewness and kurtosis of only 0.018 and –0.025 respectively. The test of normality does not permit rejection of the null hypothesis (that the distribution is normal) as its probability is 0.2.

Duration correlates highly with client cost (0.729 – Table 7) and cost (0.624 - Table 3), storeys above ground (0.606) and envelope (0.520). Also, duration correlates very significantly with most measures of cost, controlling for GIFA (Table 5). In comparison, Elhag and Boussabaine (2001) found duration to correlate significantly with the number of storeys (0.770), GIFA (0.660), tender price (0.600), frame (0.450) and foundations (0.400).

Examination of the Pearson correlations between duration and costs indicate that the untransformed duration is generally the most appropriate to use.

Procurement strategy

With the exception of three projects, the procurement strategies for the data collected are divided between traditional and design & build.

Analysis of cost functions suggests that, once building function and GIFA are taken into account, there is a significant difference between procurement strategies. Design & build is 10.98% cheaper than traditional, and management is 25.79% more expensive than traditional. Likewise, for client cost, design & build is 56.39% cheaper than traditional, and management is 14.8% more expensive than traditional. However, there are only three management projects in the data set, so these values should be treated with caution. The existence of significant differences in project cost between design & build and traditional procurement strategies was assessed by a t test and Mann-Whitney/Wilcoxon test which suggests that, even once building function and GIFA are taken into account, there is a significant difference between these two procurement routes.

Purpose

With the exception of one project, the project purpose categories for the data collected are divided between speculative and owner occupied.

The difference between the groups is best assessed using a T test and Mann-Whitney/Wilcoxon test. An analysis of variances is not valid because there is only one PFI project. The variation in the cost of the building by purpose type appears to be significant. However, when the significant variation in GIFA is taken into account there are no significant variations in cost (cost/m^2). Conversely no significant variation in raw client cost is observed, but significant variations exist in the client cost/m^2. This suggests that the project purpose is

not a significant driver of the cost of construction. However, it appears to be a significant driver of client cost.

Tendering strategy

As shown in Table 10, five tendering strategies were identified for the data collected. The significance of cost variations among these five categories is determined using analysis of variance, which shows that there is a significant difference in the construction costs between different tendering strategies, but no significant differences in the client cost. Therefore the means (of costs) have been used for ordering the variable. The order of the variables is consistent between these two cost measures, apart from partnering and selected competitive. However, the difference between the two tendering strategy types is not significant for either cost measure. Therefore, the order obtained by the (cost/m^2)/function cost measure is used, as the influence of function has been removed.

Independent site related variables

Descriptive statistics for site related variables are shown in Table 12, plus test for differenves t 13 all of which are categorised so that there is an obvious relationship between the category and the influence of that variable on the output of the model.

<<<< Insert Tables 12 & 13 about here >>>>

Site access

It is expected that the cost of the building will become more expensive as the site access becomes more restricted. An analysis of variance test shows that these differences are significant and may be attributed to the expected relationship.

Topography

While it may be expected that costs would increase as topography becomes more severe, this is not shown to be the case. Analysis of variance, however, shows that the differences in costs due to topography are not significant.

Type of location

An analysis of variance enables the null hypothesis to be accepted which shows that the cost, GIFA and function vary significantly across the different location types. However, while there is a significant difference in client cost, it is likely that this arises from the correlation between client cost and GIFA. The fact that there is not a significant difference for client costs/m^2 supports this. The hypothesis that the significant variations of cost/m^2 arise from variations of function among the location groups was tested. This was done by expressing the cost/m^2 as a proportion of the building function (the BCIS mean tender price for that function). By performing an analysis of variances on this variable, it can be shown that, when differences in the cost of different building function are taken into account, there appears to be no significant cost difference between different location types.

Site nature

The distribution of this variable was assessed using a t test. Significant differences are found in respect of cost/m^2. However, when building function is taken into account the difference is no longer significant.

Independent building definition variables

Descriptive statistics for the construction cost of building definition variables are given in Table 14, while further descriptive statistics for the categorical building function related variables are shown in Table 15.

<<<< Insert tables 14 & 15 about here >>>>

Envelope and Wall-to-floor ratio

Envelope is highly positively skewed; most of the projects appear in the 0-1000 m^2 bracket but there are a small number of very large buildings in the data.

The envelope can also be expressed in two other ways: the log of the envelope and the wall-to-floor ratio (where the envelope is divided by the GIFA). The log of the envelope has very low values of skewness and kurtosis. A test of normality does not permit rejection of the null hypothesis, indicating that the distribution is approximately normal. The most appropriate of the three representations suggested here is addressed by considering the Spearman and Pearson correlations (Tables 16 and 17 respectively) which show that many of the correlations are significant at $p = 0.001$. This suggests that the log of the envelope is the most appropriate variable to use as a predictor of cost.

Envelope correlates well with GIFA, cost, client cost, height and duration (see Table 16 and Table 18). Table 18 shows Spearman correlation coefficients between building definition variables and other variables with values above 0.4. Quality, shape, storeys below ground and structural units have no coefficients above 0.4 and the coefficient between quality and mechanical installations is 0.428.

<<<< Insert tables 16, 17 & 18 about here >>>>

Function

Function was represented by the median cost/m^2 from the broad function subgroups in the BCIS database, permitting each function to be represented as a cost value, which could then be used in correlation and regression analyses. The function variable was approximately normal and correlates well with cost/m^2, floor finishes and mechanical installations (see Tables 16 and 18). The high correlation between function and cost/m^2 is due to the way that function has been represented.

GIFA (Gross internal floor area)

GIFA is the most significant cost-influencing variable, as it represents the size of the building. Therefore, its spread is important in determining the spread of the data. GIFA has high skewness and kurtosis but the log of the GIFA is more normally distributed; a test of normality shows that the log of GIFA does not deviate from the normal distribution significantly.

GIFA correlates strongly with the following variables: envelope, cost, client cost, height, wall-to-floor ratio and lifts. In comparison, Elhag and Boussabaine (2001) found GIFA to correlate significantly with the tender price (0.940), duration (0.660) and number of storeys (0.480).

The Pearson correlation coefficients for the factored and unfactored GIFA with the major cost variables are shown in Table 17 in order to permit the most appropriate representation of floor area. This suggests that the log of the GIFA is the most appropriate representation for a

model, which predicts the log of the cost. However, when predicting the cost/m^2 the correlation is much lower, and there appears to be some disparity between the effect of the factored and unfactored variable as one is negative and the other positive. It is therefore recommended that both representations be considered for inclusion in any models to be developed and the more significant of the two selected.

Height

Height has high skewness and kurtosis but correlates significantly with envelope, GIFA, cost, client cost, storeys above ground, upper floor type, stair types and lifts. The correlation between height and storeys above ground stands to reason. The correlations with stair types and upper floors suggest that more expensive options will tend to be employed, as the building gets higher. The correlation with cost, client cost and GIFA partly arises out of the correlation with the envelope. The log of height is approximately normally distributed, with a skewness and kurtosis of 0.474 and 0.070. However, a test of normality does cause the null hypothesis to be rejected.

In order to determine whether the factored or unfactored variable should be used, the Pearson correlation coefficients on the major cost variables are presented in Table 17. This suggests that the factored variable (the log) is only appropriate for models which predict the log of the raw cost.

Quality

Quality is an ordinal 5-point scale, as shown in Table 15. Nearly half the projects are defined as being "medium" quality. Only around 20 projects each are described as being of low or low-medium quality. The former is simply a measure of the fact that very few clients/QSs

would admit that the quality of a building is low. Quality correlates strongly with no variables, but has smaller (statistically significant) correlations with many cost, finish related and mechanical and electrical (M&E) related variables, which is consistent with the hypothesis that there is an overall level of specification, which these variables will tend to follow. The fact that quality correlates with cost is a function of the fact that buildings of a higher specification will cost more.

Shape complexity

The cost/m^2 results suggest that the order is correct, confirming Ashworth's (2004) assertion that the more complex the shape, the higher the cost. However, there is only a small difference between low and medium category buildings. Further, when variations in function are taken into account, the results suggest that buildings of medium shape complexity are actually less expensive than those of low complexity. This was assessed by a T test, as shown in Table 19.

<<<< Insert table 19 about here >>>>

This suggests that the differences in cost could have arisen by chance. It is possible that the increase in cost one intuitively expects when moving from a building of low shape complexity to one of medium complexity is small in comparison to the standard error of the mean costs. Thus random variations in the cost could be higher than the cost difference itself, and mask the true effect. However, it is also possible that the apparent variation arises from correlations with other cost significant variations. Perhaps the best suggestion is that the three possible orders for the variables are tested for inclusion within a stepwise regression. The order that gives the highest value of T for inclusion would then be the most appropriate order.

Storeys above ground

The high kurtosis of storeys above ground arises from the fact that over half of the projects are less than 2 storeys high and the representation of a higher number of storeys is quite low, possibly lending itself to representation in logarithmic form. The log of storeys above ground is better spread out, with kurtosis and skew of 0.900 and 0.367 respectively. However, there are still two very large peaks at the values corresponding to one and two storeys, and this makes the distribution fail the test of normality. The most suitable of the raw and logarithmic values to use is estimated from the Pearson correlation with the cost variables (Table 20) which suggests that the unconverted value should be used for all cost values except the log of cost/m^2.

<<<< Insert table 20 about here >>>>

Those variables with a strong correlation with storeys above ground are client cost, duration, height, stair type and upper floors. Elhag and Boussabaine (2001) found number of storeys to correlate significantly with duration (0.770), working space (0.510), GIFA (0.480), foundations (0.450) and frame (0.450). The correlations with duration could be caused by the fact that it is not possible to commence work on the next storey until the current one is sufficiently completed. Correlation with height is to be expected. The correlation with upper floors and stair type suggests that particular stair and floor solutions are associated with buildings that have more storeys.

Storeys below ground

230 of the projects analysed had no basement storeys and so no strong correlations (above 0.5) were observed; the highest correlation was with site access (0.426).

Structural units

The distribution of structural units is highly skewed, as only 21% of the projects have a value of more than 1. This could make it suitable for expression as a logarithm, however, taking the log does not yield a substantial improvement in the distribution.

Structural units does not correlate well with other descriptor variables. The highest correlation is with envelope (0.362).

CONCLUSIONS

This, the first of two papers, seeks to redress the omission in recent literature on the influence of project strategic, site related and design related variables on the cost of construction, by presenting the results of an investigation into the influence of 41 independent variables on both construction cost and client on cost. Data were collected from 286 construction projects in the United Kingdom and analysed to determine the relationships that exist between the dependent and independent variables.

The analysis confirms the strong relationship between construction cost and client cost and between those two measures of cost and GIFA, indicating that the most appropriate output for modelling purposes is cost/m^2 or log of cost/m^2, both of which are more closely aligned to the normal distribution. In addition to client cost and GIFA, construction cost is well correlated

with envelope, height, upper floors, lifts and duration. As expected, client cost correlates well with those variables which correlate well with construction cost.

In respect of project strategic variables, analysis suggests that, once building function and GIFA are taken into account, there is a significant difference between procurement strategies in terms of both construction and client cost; design & build is significantly cheaper than traditional. In respect of tendering strategy, where five categories were observed, there is a significant difference in the construction costs between different tendering strategies, but no significant differences in the client cost. Therefore the means (of costs) have been used for ordering the variable. Conversely, once GIFA is taken into account, there are no significant variations in construction cost in respect of purpose (speculative or owner occupied) although there are significant variations in client cost/m^2.

In respect of site related variables, it was found that as site access became more restricted there was a significant increase in cost. Brownfield sites were also significantly more costly than greenfield sites and centrally urban located sites were more costly than sites located rurally. However, when building function is taken into account these differences are no longer significant. Conversely, an increase in the severity of the topography of the site led to a decrease in costs, although the cost differences were not significant.

In respect of the building function variables, GIFA is the most significant cost influencing variable and it correlates well with envelope, cost, client cost, height and wall-to-floor ratio an lifts. The log of GIFA does not deviate significantly from the normal distribution and it is suggested that this is the most appropriate way to represent GIFA in any modelling. The same observation can be made in respect of height. Quality correlates strongly with no variables,

but has smaller (but still statistically significant) correlations with many cost, finish related and services installations variables, which is consistent with the hypothesis that there is an overall level of specification, which these variables will tend to follow.

Overall, the analyses described in this paper have established the core relationships which exist within the data and confirmed many of the relationships that had been anticipated from the literature. It also established the ordinal sequence of several nominal variables. These data, therefore, can be confidently used to develop models of the total cost of construction.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the Engineering and Physical Sciences Research Council (EPSRC), who funded the research through two grants; our industrial collaborators: Paul Moore - EC Harris, the late Chris Powell – formerly with Tweeds (then Faithful and Gould), Alun Williams – Symonds, and Joe Martin – Building Cost Information Service (BCIS); and the contribution made by the research assistants: Mick Gregory and Adam Hickson and by Dr Roy Duff, former senior lecturer, Manchester Centre for Civil and Construction Engineering, UMIST.

REFERENCES

Ashworth, A. (2004) *Cost Studies of Buildings.* 4th ed. Pearson, Prentice Hall, Harlow, Essex, UK.

Braby, R. H. (1975) Costs of high-rise buildings. *The Building Economist*, 14(2), 84-86.

Brandon, P. S., Baseden, A., Hamilton, I. W. and Stockley, J. E. (1988) *Applications of Expert Systems to Quantity Surveying: the Strategic Planning of Construction Projects*, Surveyors Publications, RICS, London

Brandon, P. S. (1990) The development of an expert system for the strategic planning of construction projects. Construction Management and Economics, 8(3), 285-300.

Corbett, P. and Rowley, R. (1999) *The use of BCIS elemental cost data by quantity surveyors as part of cost planning techniques: the practitioners' perspective*. 15th Annual Conference, ARCOM, Liverpool John Moores University. 465-472.

Elhag, T. and Boussabaine, A. H. (2001) Tender price estimation using artificial neural networks. *Journal of Financial Management of Property and Construction*, 6(3), 193-208.

Elhag, T. and Boussabaine, A. H. (2002) Tender price estimation using artificial neural networks II: modelling. *Journal of Financial Management of Property and Construction*, 7(1), 49-64.

Emsley, M. W., Lowe D. J., Duff, A. R., Harding, A. and Hickson, A. (2002) Development of neural networks to predict total construction costs. *Construction Management and Economics,* **20**, 465-472

Ferry, D. J., Brandon, P. S. and Ferry, J. D. (1999) *Cost Planning of Buildings* 7th Edition, Blackwell Science Ltd, Oxford.

Flanagan, R. and Norman, G. (1978) The relationship between construction price and height, *Chartered Surveyor B and QS Quarterly*, Summer, 69-71

Günaydm, H. M. and Goğan, S. Z. (2004) A neural network approach for early cost estimation of structural systems of buildings. *International Journal of Project Management*, 22, 595-602.

Harding, A., Lowe, D. J., Emsley, M. W., Hickson, A. and Duff, A. R. (1999) *Implementation of a neural network for the comparison of the cost of different procurement approaches.* 15th Annual Conference, ARCOM, Liverpool John Moores University. 763-772.

Harding, A., Lowe, D. J., Emsley, M. W., Hickson, A. and Duff, A. R. (2000) *Implementation of a neural network model for the comparison of the cost of different procurement approaches.* Information and Communication in Construction Procurement, CIB W92 Construction Procurement System Symposium, Santiago, Chile. 269-280.

Jagger, D., Ross, A., Smith, J. and Love, P. (2002) *Building Design Cost Management*. Blackwell Publishing, Oxford.

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

Kim, G-H., Yoon, J-E., An, S-H., Cho, H-H. and Kang, K-I. (2004b) Neural network model

incorporating a genetic algorithm in estimating construction costs. *Building and Environment*, 39, 1333-1340.

Lowe, D. J., Emsley, M.W. and Harding, A. (2006a) Predicting construction cost using multiple regression techniques. *Journal of Construction Engineering and Management*, American Society of Civil Engineers, 132 (7), 750-758.

Lowe, D. J., Emsley, M.W. and Harding, A. (2006b) The relationship between total construction cost and design related variables. Submitted to the *Journal of Financial Management of Property and Construction*.

McCaffer, R. (1975) Some examples of the use of regression analysis as an estimating tool. *Quantity Surveyor,* December, 81-86.

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

Morton, R. and Jagger, D. (1995) *Design and the Economics of Building*, E & F N Spon, London.

Newton, S. (1991) An agenda for cost modelling research. *Construction Management and Economics,* 9(2), 97-112.

Pegg, I. (1984) *The effect of location and other measurable parameters on tender levels,* Cost

Study F33, BCIS, 1983-4-219-5 to 17, Building Cost Information Service, RICS, London.

Picken, D. H. and Ilozor, B. D. (2003) Height and construction costs of buildings in Hong Kong. *Construction Management and Economics,* 21(2), 107-111.

Seeley, I. H. (1995) *Building Economics*, 4th Edition, Palgrave, London

Skitmore, M. and Patchell, B. (1995) Developments in Contract Price Forecasting and Bidding Techniques, in P. S. Brandon (ed) *Quantity Surveying Techniques: New Directions*, Blackwell Scientific, Oxford, pp 75-120.

Swaffield, L. M. and Pasquire, C. L. (1999) Examination of relationships between building form and function, and the cost of mechanical and electrical service. *Construction Management and Economics,* 17(4), 483-492.

Tan, W. (1999) Construction cost and building height. *Construction Management and Economics,* 17(2), 129-132.

Tarassenko, L. (1998). *A guide to neural computing applications*. Arnold, London.

Trost, S. M. and Oberlender, G. D. (2003) Predicting Accuracy of Early Cost Estimates Using Factor Analysis and Multivariate Regression. *Journal of Construction Engineering and Management,* 129(2), 198-204.

Wilderness Group (1964) *An investigation into building cost relationships of the following design variables: storey height, floor loading, column spacing, number of storeys*, Report to the Royal Institution of Chartered Surveyors, RICS, London

Wing, C. K. (1999) On the issue of plan shape complexity: plan shape indices revisited. *Construction Management and Economics,* **17**(4), 473-482.

Table 1: Classification of input variables

Project strategic variables								
Contract form	CFORM	(O)	Procurement strategy	PROC	(O)	Tendering strategy	TEND	(O)
Duration	DUR	(S/weeks)	Purpose	PURP	(O)			
Site related variables								
Site access	ACCS	(O)	Type of location	LOC	(O)			
Topography	TOP	(O)	Nature of site	SITE	(N)			
Design related variables								
Building definition			Structure			Finishes and openings		
Envelope	ENVLP	(S/m^2)	External walls	EXWL	(S/£)	Ceiling finishes	CEILF	(S/£)
Function	FUNC	(S/£)	Frame	FRAME	(O)	External doors	EXDR	(S/£)
GIFA	GIFA	(S/m^2)	Piling	PILING	(N)	Floor finishes	FLRF	(S/£)
Height	HEIGHT	(S/m)	Roof construction	ROOF	(O)	Internal doors	INDR	(S/£)
Quality*	QUAL	(O)	Roof profile	RFPRO	(O)	Internal walls	INWL	(S/£)
Shape complexity	SHAPE	(O)	Stair types	STAIRS	(O)	Internal wall finishes	INWLF	(S/£)
Storeys above ground	STABV	(S/number)	Substructure	SUB	(O)	Roof finishes	ROOFF	(S/£)
Storeys below ground	STBL	(S/number)	Upper floors	UPFL	(O)	Windows	WINS	(S/£)
Structural units	UNITS	(S/number)				Services installations		
Wall-to-floor ratio	WALFL	(S/ratio)				Air conditioning	AC	(S/£)
						Electrical installations	ELEC	(S/£)
						No. lifts	LIFTS	(S/number)
						Mechanical installations	MECH	(S/£)
						Protective installations	PROT	(S/£)
						Special installations	SPEC	(S/£)

Note: N = nominal, (e.g., piling: 0 = no piling, 1 = piling); O = ordinal (e.g., shape complexity: 0 = low, 1 = medium, 2 = high); and S = scale (e.g., external doors: a proportional scale based on relative costs).

* Note that in previous publications *Quality* has been classified as a project strategic variable. Here it has been re-classified as a design related variable.

Lowe, D J, Emsley, M W and Harding, A (2006) Relationships between total construction cost and project strategic, site related and building definition variables. *Journal of Financial Management of Property and Construction*, **11**(3), 165-180, eScholarID:1g121 | DOI:10.1108/13664380680001087

Table 2: Descriptive statistics for construction cost variables

	COST	***LN*COST**	**COSTM^2**	***LN*COSTM^2**
N	286	286	286	286
Mean (£)	1268027.97	13.27	381.58	5.84
Median (£)	564403.23	13.24	358.37	5.88
Std. deviation (£)	2018424.59	1.23	172.30	0.47
Skewness	3.68	0.27	0.91	-0.38
Kurtosis	17.75	-0.39	0.93	0.30
Minimum (£)	35918.69	10.49	56.45	4.03
Maximum (£)	15792339.59	16.58	1026.51	6.93

Table 3: Spearman correlation coefficients: Construction cost

COST/*LN*COST		**COSTM^2/*LN*COSTM^2**	
GIFA	0.930	CCOSTM^2	0.708
CCOST	0.872	FUNC	0.698
ENVLP	0.863	MECH	0.592
HEIGHT	0.698	CCOST	0.573
UPFL	0.651	FLRF	0.572
LIFTS	0.631	INWLF	0.570
DUR	0.624		
STRS	0.580		
PROT	0.579		
UPFL	0.566		
STAIRS	0.535		
WALFL	-0.529		
STABV	0.511		

Table 4: Pearson correlation coefficients: Construction cost

COST		***LN*COST**		**COSTM^2**		***LN*COSTM^2**	
GIFA	0.865	*LN*GIFA	0.925	CCOSTM^2	0.755	CCOSTM^2	0.713
CCOST	0.859	*LN*CCOST	0.880	*LN*CCOSTM^2	0.684	LNFUNC	0.701
*LN*CCOST	0.749	*LN*ENVLP	0.856	FUNC	0.654	*LN*CCOSTM^2	0.696
LIFTS	0.726	GIFA	0.790	*LN*FUNC	0.647	FUNC	0.691
*LN*GIFA	0.687	ENVLP	0.747	*LN*CCOST	0.589	MECH	0.579
ENVLP	0.678	*LN*HEIGHT	0.721	MECH	0.550	*LN*FLRF	0.575
*LN*LIFTS	0.673	CCOST	0.712	CCOST	0.541	*LN*CCOST	0.564
HEIGHT	0.629	HEIGHT	0.671	*LN*INWLF	0.517	*LN*INWLF	0.540
DUR	0.628	*LN*DUR	0.659	*LN*FLRF	0.512	CEILF	0.521
*LN*ENVLP	0.620	DUR	0.651	FLRF	0.504	FLRF	0.508
PROT	0.600	LIFTS	0.630	INWLF	0.500	*LN*ICEILF	0.505
*LN*HEIGHT	0.598	UPFL	0.628				
STABV	0.565	PROTECT	0.587				
*LN*EXWL	0.549	*LN*STBV	0.553				
*LN*DUR	0.540	STAIRS	0.550				
EXWL	0.531	STABV	0.548				
UPFL	0.512	*LN*WALFL	-0.519				
*LN*STABV	0.507	WALFL	-0.505				

Table 5: Partial correlation coefficients controlling for GIFA: Construction cost

COST		***LN*COST**		**COSTM^2**		***LN*COSTM^2**	
CCOST	0.772	*LN*CCOST	0.813	CCOST	0.842	*LN*CCOSTM^2	0.792
*LN*CCOSTM^2	0.693	*LN*ENVLP	0.636	*LN*CCOSTM^2	0.810	CCOSTM^2	0.774
LIFTS	0.560	*LN*DUR	0.592	COSTFUNC	0.696	*LN*FUNC	0.700
*LN*CEIF	0.551	DUR	0.522	FUNC	0.651	FUNC	0.690
CCOSTM^2	0.536	*LN*HEIGHT	0.513	*LN*CCOST	0.646	*LN*CCOST	0.661
EXWL	0.526	*LN*STABV	0.505	*LN*FUNC	0.644	COSTFUNC	0.649
*LN*LIFTS	0.519			CCOST	0.585	MECH	0.585
*LN*EXWL	0.512			MECH	0.547	*LN*FLRF	0.575
STABV	0.510			*LN*INWLF	0.517	CCOST	0.559
DUR	0.508			*LN*FLRF	0.508	*LN*INWLF	0.539
						CEILF	0.521
						*LN*CEILF	0.520
						FLRF	0.507
						DUR	0.505

Table 6: Descriptive statistics for client cost variables

	CCOST	**LNCCOST**	**CCOSTM^2**	**LNCCOSTM^2**
N	67	67	67	67
Mean (£)	83287.25	10.80	40.58	3.46
Median (£)	45363.53	10.72	35.24	3.56
Std. deviation (£)	113147.49	0.99	26.38	0.75
Skewness	3.27	0.44	0.81	-0.60
Kurtosis	11.86	0.03	-0.07	0.17
Minimum (£)	5863.16	8.68	3.57	1.27
Maximum (£)	637497.78	13.37	109.45	4.70

Table 7: Spearman correlation coefficients: Client cost

CCOST/*LN*CCOST		**CCOSTM^2/*LN*CCOSTM^2**	
COST	0.872	COSTM^2	0.708
DUR	0.729		
GIFA	0.719		
ENVLP	0.716		
HEIGHT	0.676		
STABV	0.636		
PROT	0.623		
LIFTS	0.610		
COSTM^2	0.573		
MECH	0.544		
AC	0.534		
UPFL	0.519		
STAIRS	0.513		
INDR	0.506		

Table 8: Pearson correlation coefficients: Client cost

CCOST		**LNCCOST**		**CCOSTM^2**		**LNCCSTM^2**	
COST	0.859	*LN*COST	0.880	COSTM^2	0.755	*LN*COSTM^2	0.696
LIFTS	0.746	COST	0.749	*LN*COSTM^2	0.713	COSTM^2	0.684
*LN*COST	0.712	*LN*DUR	0.746				
PROT	0.678	DUR	0.743				
DUR	0.659	*LN*GIFA	0.722				
GIFA	0.657	*LN*HEIGHT	0.705				
HEIGHT	0.656	*LN*ENVLP	0.687				
*LN*HEIGHT	0.621	PROT	0.685				
RFPRO	0.617	HEIGHT	0.682				
*LN*STABV	0.615	*LN*STABV	0.655				
*LN*DUR	0.604	GIFA	0.651				
STABV	0.595	LIFTS	0.630				
ENVLP	0.576	ENVLP	0.602				
*LN*GIFA	0.571	STABV	0.600				
*LN*LIFTS	0.556	COSTM^2	0.589				
*LN*ENVLP	0.544	UPFL	0.584				
COSTM^2	0.541	*LN*STBL	0.579				
MECH	0.507	*LN*COSTM^2	0.564				
		MECH	0.557				
		STAIRS	0.553				
		*LN*INDR	0.526				
		EXWL	0.517				
		*LN*PROT	0.511				
		INDR	0.509				
		RFPRO	0.502				
		AC	0.500				

Table 9: Spearman correlation coefficients: Construction and client costs

	COST 286	$COSTM^2$ 286	CCOST 67	$CCOSTM^2$ 67
COST	1.000	0.334***	0.872***	-0.098
$COSTM^2$		1.000	0.573***	0.708***
CCOST			1.000	0.307*
$CCOSTM^2$				1.000

*** = p≤ 0.001 ** = p≤ 0.01 *= p≤ 0.05

Table 10: Descriptive statistics for project strategic variables

			Cost/m^2		Client Cost/m^2	(Cost/m^2)/ Function	Function
Contract form	**IFC**	31	£ 368.20	9	£45.22	0.647	£573.68
	JCT – traditional	91	£ 374.31	33	£46.24	0.645	£575.54
	JCT – contractor's design portion	23	£ 489.94	2	£63.51	0.752	£667.35
	JCT – design & build	120	£ 352.28	18	£22.35	0.595	£598.28
	G/C works	13	£ 534.08	1	£42.09	0.766	£743.23
	Other (bespoke)	8	£ 396.18	4	£53.62	0.569	£689.38
Procurement strategy	**Design & build**	130	£ 357.16	21	£24.26	0.594	£606.62
	Traditional	153	£ 400.17	44	£47.24	0.668	£599.45
	Management	3	£ 491.24	2	£65.39	0.840	£633.67
Purpose	**Speculative**	65	£368.38	45	£28.33	0.612	£603.37
	Owner occupied	220	£384.50	22	£46.57	0.642	£602.48
	PFI	1	£595.83	-	-	0.836	£713.00
Tendering strategy	**Negotiated**	46	£ 367.27	7	£30.32	0.595	£616.85
	Selected competitive (single stage)	213	£ 370.30	58	£41.43	0.631	£591.40
	Partnering	8	£ 391.37	1	£16.82	0.622	£626.25
	Open	5	£ 489.76	1	£86.87	0.688	£703.40
	Two stage	14	£ 555.92	-	-	0.838	£686.36

Table 11: Tests for differences between sub-groups of project strategic variables

	CFORM		PROC		PURP		TEND	
	F	**Chi-S**	**T**	**Z**	**t**	**Z**	**F**	**Chi-S**
COST	11.079***	64.237***	-0.718	-1.023	-2.704**	-3.673***	12.263***	18.879***
***LN*COST**	16.857***	-	0.615		-3.912***		8.186***	
COSTM^2	4.953***	25.819***	-2.155*	-1.619	0.663	-0.814	4.613***	16.726***
***LN*COSTM^2**	4.741***	-	-1.148		0.943		3.777**	
CCOST	4.809***	18.483**	1.232	-2.399*	0.190	-0.053	0.516	3.027
***LN*CCOST**	5.269***	-	2.154*		-0.263		0.601	
CCOSTM^2	2.960*	14.805*	-4.382***	-3.282***	3.372***	-2.323*	1.273	3.025
***LN*CCOSTM^2**	4.294**	-	-3.963***		2.441*		0.883	
COSTFN	2.982*	15.731**	-2.776**	-2.225*	0.918	-0.957	3.187*	11.684**
FUNC	2.965*	13.551*	0.319	-0.461	-0.031	-0.777	1.342	7.037
***LN*FUNC**	2.760*	-	0.493		0.394		1.448	
GIFA	7.135***	54.497***	0.033	-1.271	-3.013**	-4.451***	9.554***	13.079**
***LN*GIFA**	12.733***	-	1.110		-4.611**		5.796***	

Table 12: Descriptive statistics for site related variables

			Cost/m^2		Client Cost/m^2	(Cost/m^2)/ Function	Function
Site access	**Unrestricted**	186	£341.25	46	£33.01	0.609	£564.81
	Restricted	70	£427.53	16	£56.00	0.633	£672.91
	Highly restricted	29	£530.07	5	£60.88	0.821	£679.79
Topography	**Level site**	182	£376.52	41	£39.55	0.637	£594.05
	Gently sloping site	67	£382.74	17	£46.03	0.654	£586.88
	Mixed/undulating	18	£435.45	4	£27.92	0.637	£684.83
	Steeply sloping	18	£375.70	5	£40.64	0.562	£672.56
Type of location	**Central urban**	61	£449.98	11	£46.80	0.700	£654.48
	Other urban	161	£360.38	38	£42.68	0.606	£597.56
	Rural	63	£369.87	18	£32.35	0.652	£567.33
Site nature	**Greenfield**	110	£346.49	31	£31.48	0.600	£575.50
	Brownfield	175	£403.76	36	£48.42	0.659	£620.38

Table 13: Tests for differences between sub-groups of site related variables

	ACCS		TOP		LOC		SITE	
	F	**Chi-S**	**F**	**Chi-S**	**F**	**Chi-S**	**t**	**Z**
COST	9.279***	10.017***	1.521	3.527	7.200***	14.945***	-2.430*	-2.547*
***LN*COST**	6.761***		1.317		8.694***		-2.621**	
COSTM^2	20.848***	41.268***	0.642	3.874	6.384***	10.519**	-2.759**	-2.838**
***LN*COSTM^2**	21.154***		1.111		5.217***		-2.719**	
CCOST	4.810*	2.781	3.589*	5.019	9.044***	13.004**	-1.612	-1.283
LNCCOST	2.513		2.241		9.245***		-1.520	
CCOSTM^2	7.264***	11.735**	0.558	1.451	1.314	1.444	-2.747**	-2.352*
***LN*CCSTM^2**	6.942**		0.264		0.267		-2.290*	
COSTFN	11.203***	16.488***	0.741	2.685	3.839*	4.589	-2.101*	-1.867
FUNC	11.857***	21.288***	2.273	7.027	3.523*	8.199*	-1.965*	-2.302*
***LN*FUNC**	12.945***		2.162		3.468*		-1.903	
GIFA	2.547	5.192	1.490	3.749	3.492*	10.844**	-1.780	-1.651
***LN*GIFA**	3.089*		1.275		5.511**		-1.675	

Table 14: Descriptive statistics for the building definition variables

	ENVLP	*LN*ENVLP	WALFL	FUNC	*LN*FUNC	FUNC (group)	GIFA	*LN*GIFA
Mean	1724.693	7.001	0.715	603.066	6.350	3.150	3160.057	7.433
Median	1090.800	6.995	0.670	593.000	6.385	4.000	1612.500	7.386
Std. Dev	1741.861	0.986	0.292	188.376	0.3289	2.313	3965.201	1.154
Skewness	2.255	-0.076	0.770	0.185	-0.285	-0.167	2.631	-0.008
Kurtosis	7.395	-0.695	0.598	-0.851	-1.008	-1.457	8.522	-0.560
Min.	130.00	4.87	0.21	318.00	5.76	0	97.00	4.57
Max.	11928.90	9.39	1.86	1126.00	7.03	8	26219.00	10.17
	HEIGHT	***LN*HEIGHT**	**QUAL**	**STABV**	***LN*STABV**	**STBL**	***LN*STBL**	**UNITS**	***LN*UNITS**
Mean	7.444	1.840	1.205	2.070	.563	0.101	0.137	1.927	.841
Median	6.000	1.792	1.000	1.810	0.593	0.000	0.000	1.000	.693
Std. Dev	5.009	0.556	0.583	1.439	0.537	0.357	0.312	3.034	.576
Skewness	2.235	0.474	-0.120	2.485	0.829	3.777	1.126	3.792	1.684
Kurtosis	6.248	.070	-0.550	7.805	0.178	14.447	-0.264	15.756	3.594
Min.	2.100	0.74	0.000	1.000	0.000	0.000	-0.268	0.000	0.000
Max.	33.50	3.51	2.0	10.00	2.30	2.00	0.693	20.00	3.04

Table 15: Descriptive statistics for categorical building definition variables

			Cost/m^2		Client Cost/m^2	(Cost/m^2)/ Function	Function
Building function	**Industrial**	66	£234.47	23	24.64	0.588	£404.50
	Admin/Commercial	69	£431.16	15	37.48	0.681	£652.91
	Educational	28	£448.93	11	66.08	0.600	£748.11
	Health/welfare	25	£532.62	3	41.44	0.668	£798.08
	Recreational	34	£434.41	5	66.46	0.604	£727.74
	Religious	3	£666.82	0	-	0.635	£1031.67
	Residential	71	£357.34	10	40.63	0.642	£556.17
Quality	**Low quality**	21	£207.19	5	£28.10	0.527	£396.76
	Low-medium quality	21	£262.29	1	£16.82	0.560	£475.57
	Medium quality	139	£369.32	39	£40.75	0.610	£608.89
	Medium-high quality	30	£501.75	4	£40.67	0.741	£707.43
	High quality	75	£438.45	18	£44.98	0.695	£643.99
Shape complexity	**Low**	155	£358.15	41	£37.82	0.631	£552.55
	Medium	89	£385.89	16	£44.16	0.609	£639.49
	High	36	£500.53	7	£37.05	0.717	£717.19

Table 16: Spearman correlation coefficients between construction and client costs and building definition variables

		COST	COSTM^2	CCOST	CCOSTM^2
ENVLP	**277**	0.863***	-0.034	0.716***	-0.382**
FUNC	**286**	0.314***	0.698***	0.457***	0.423***
FUNCGP	**286**	-0.264***	-0.392***	-0.395***	-0.338**
GIFA	**286**	0.930***	0.003	0.719***	-0.389***
HEIGHT	**280**	0.698***	0.144*	0.676***	-0.125
QUAL	**286**	0.330***	0.402***	0.332***	0.181
SHAPE	**280**	0.239***	0.284***	0.262*	0.020
STABV	**286**	0.511***	0.392***	0.636***	0.190
STBL	**286**	0.357***	0.322***	0.310*	0.130
UNITS	**286**	0.093	-0.217***	0.103	-0.305*
WLFLR	**277**	-0.529***	-0.092	-0.450***	0.240

*** = p≤ 0.001 ** = p≤ 0.01 *= p≤ 0.05

Table 17: Pearson correlation coefficients between construction and client costs and building definition variables

		COST	*LN* COST	COSTM^2	*LN* COSTM^2	CCOST	*LN* CCOST	CCOSTM^2	*LN* CCOSTM^2
ENVLP	**277**	0.678***	0.747***	0.009	-0.009	0.576***	0.602***	-0.264*	-0.396***
***LN*ENVLP**	**277**	0.620***	0.856***	-0.045	-0.049	0.544***	0.687***	-0.348**	-0.438***
FUNC	**286**	0.291***	0.308***	0.654***	0.691***	0.377**	0.441***	0.463***	0.468***
***LN*FUNC**	**286**	0.280***	0.294***	0.647***	0.701***	0.364**	0.427***	0.462***	0.477***
FUNCGP	**286**	-0.217***	-0.274***	-0.321***	-0.382***	-0.318**	-0.390***	-0.353**	-0.336**
GIFA	**286**	0.865***	0.790***	0.091	0.050	0.657***	0.651***	-0.274*	-0.368**
***LN*GIFA**	**286**	0.687***	0.925***	-0.012	-0.040	0.571***	0.722***	-0.363**	-0.423***
HEIGHT	**280**	0.629***	0.671***	0.297***	0.270***	0.656***	0.680***	0.032	0.016
***LN*HEIGHT**	**280**	0.598***	0.721***	0.197***	0.172**	0.621***	0.705***	-0.041	-0.106
QUAL	**286**	0.242***	0.324***	0.384***	0.420***	0.217	0.314**	0.156	0.176
STABV	**280**	0.565***	0.548***	0.405***	0.410***	0.595***	0.600***	0.153	0.166
***LN*STABV**	**286**	0.507***	0.553***	0.400***	0.433***	0.615***	0.655***	0.213	0.195
STBL	**286**	0.495***	0.409***	0.353***	0.307***	0.331**	0.362**	0.105	0.141
***LN*STBL**	**24**	0.340	0.422*	0.209	0.210	0.407	0.579	-0.642	-0.608
SHAPE	**280**	0.142*	0.241***	0.270***	0.296***	0.426***	0.349**	0.040	-0.012
UNITS	**286**	-0.046	0.025	-0.187***	-0.156**	-0.062	0.006	-0.146	-0.151
***LN*UNITS**	**286**	-0.051	0.002	-0.211***	-0.188**	-0.129	-0.093	-0.162	-0.209
WLFLR	**277**	-0.410***	-0.505***	-0.067	-0.026	-0.256*	-0.357**	0.252*	0.193
***LN*WLFLR**	**277**	-0.467***	-0.519***	-0.076	-0.030	-0.285*	-0.404***	0.246	0.181

*** = p≤ 0.001 ** = p≤ 0.01 *= p≤ 0.05

Table 18: Spearman correlation coefficients between building definition variables and other variables

ENVLP		**FUNC**		**GIFA**	
GIFA	0.933	FLRF	0.560	ENVLP	0.933
HEIGHT	0.709	MECH	0.514	HEIGHT	0.709
DUR	0.520	INWLF	0.487	UPFL	0.551
STABV	0.457	LIFTS	0.439	WLFLR	-0.547
STAIRS	0.454	UPFL	0.439	LIFTS	0.537
UPFL	0.454	CEILF	0.421	STAIRS	0.507
LIFTS	0.443	STABV	0.408	DUR	0.481
HEIGHT		**STABV**		**WLFLR**	
ENVLP	0.709	HEIGHT	0.662	GIFA	-0.547
GIFA	0.709	DUR	0.606	LIFTS	-0.445
UPFL	0.673	STAIRS	0.633	UPFL	-0.436
STABV	0.662	LIFTS	0.530	PROT	-0.421
STAIRS	0.601	UPFL	0.516		
LIFTS	0.577	INWLF	0.468		
PROT	0.432	ENVLP	0.457		
AC	0.416	MECH	0.454		
		FUNC	0.408		
		INDR	0.400		

Table 19: T-test for low/medium shape complexity

	Cost/m^2	**Client Cost/m^2**	**(Cost/m^2)/ Function**	**Function**
	286	**67**	**286**	**286**
Low/medium	-1.723	-0.745	0.827	-3.522***

Table 20: Pearson correlation coefficients for storeys above ground

	Raw cost	**Log of cost**	**Raw cost/m^2**	**Log of cost/m^2**	**(Cost/m^2)/ function**
Raw value	0.565***	0.548***	0.405***	0.410***	0.233***
Log value	0.507***	0.553***	0.400***	0.433***	0.182**