MODELLING PROBABILITY DISTRIBUTIONS OF FACILITIES MANAGEMENT COSTS FOR AN NHS ACUTE CARE HOSPITAL BUILDING

Richard J. Kirkham and A. Halim Boussabaine

Construction Cost Engineering Research Unit, School of Architecture and Building Engineering, University of Liverpool, L69 3BX, UK

Whole Life Cycle Costing (WLCC) provides a valuable insight into the economic efficiency of constructed facilities. It has however, been criticized by practitioners and academics alike for producing forecasts of a high-risk nature, a consequence of the large number of assumptions inherent in the modelling process. Most WLCC systems use deterministic assumptions but a stochastic approach to modelling the assumptions will achieve a better representation of the likely costs. Monte Carlo and Latin Hypercube simulation can be used to create forecasts of the whole life costs but the accuracy of these are strongly correlated to the quality of the input probability distributions, hence the purpose of this paper is to present a methodology for the definition of probability distributions that best represent the facilities management costs in an acute care NHS hospital building

The data used in this research were obtained from National Health Service (NHS) Estates on the Facilities Management (FM) costs of over 450 acute care NHS Trusts in England and Wales. The data was analysed to obtain the parameters of the theoretical distributions that best describe the FM costs for a local NHS Acute Care teaching hospital building. The distributions were then validated using various goodness-of-fit techniques. The result of this work might then be used as stochastic assumptions in the modelling of WLCC. The paper also discusses some issues of accuracy in distribution fitting, particularly the class interval rule and its effect on the quality of the results, and the selection of an appropriate goodness-of-fit test.

Keywords: facilities management, hospital buildings, National Health Service acute care, probability density functions, whole life costing.

INTRODUCTION

WLCC is widely recognized amongst practitioners and academics as a valuable tool in assessing the economic efficiency of constructed facilities. It can be used as a means of comparing options and their associated costs and incomes over a period of time (CBPP 1998), or as a tool for assessing the long terms costs of ownership in existing buildings through stochastic modelling and key performance indicators (Kirkham and Boussabaine 2000a).

WLCC though is to a significant extent, dependent on assumptions about the future costs of operating and maintaining the building and its environment. It has been widely noted that concerns about using a WLCC approach are based mainly on the risky nature of the assumptions on which the forecasts are modelled (BRE 1999, Jovanovic 1999, Edwards and Bowen 1998). Whilst forecasting of future costs is to some extent an inexact science, this should not dissuade analysts and managers from attempting to apply WLCC principles (Woodward 1997).

Kirkham, R J and Boussabaine, A H (2001) Modelling probability distributions of facilities management costs for an NHS acute care hospital building. *In:* Akintoye, A (Ed.), *17th Annual ARCOM Conference*, 5-7 September 2001, University of Salford. Association of Researchers in Construction Management, Vol. 1, 401-10.

As one of the leading providers of acute healthcare services in Europe, the United Kingdom NHS operates and manages a complex property portfolio of high occupancy buildings and establishments. Approximately £33.3bn of UK government spending is apportioned to the NHS, and some 22% of this accounts for estate management and capital investment (NHS Estates 1999a). The costs of maintaining and operating these buildings constitutes a significant proportion of total NHS expenditure on the estate, and in particular, the FM costs represent an integral part of this expenditure. As part of Audit Commission guidance on NHS FM expenditure (Audit Commission 2000), a large proportion of the FM work is now outsourced to external private contractors (NHS Estates 1996, NHS Estates 1998, NHS Estates 1999b), and equipping professionals with the ability to monitor the cost effectiveness of these services should be an integral part of any NHS Trusts financial management programme.

NHS Estates collate annual data on FM costs from NHS trusts as part of the Trust Financial Proformas (TFP) returns (NHS Estates 1997), this gives stakeholders a snapshot of the variance of FM costs through the entire NHS estate. This is generally used to create simple benchmarks whereby NHS trusts can gauge the economic efficiency of their FM services against national averages.

However, the nature of this data poses problems for use in WLCC exercises. As the data is collated at trust level, it represents total spending on FM costs for all buildings within the Trust’s estate, not for individual sites that make up the entire trust property portfolio. Most acute care trusts encompass several buildings within the estate and hence the data in its raw form cannot be used to monitor FM costs for individual buildings. To facilitate an accurate simulation of FM costs, the data must be transformed and analysed to reflect the cost of individual buildings within a Trust estate portfolio, not for the estate as whole. This paper proposes a methodology for addressing this issue.

RESEARCH METHODOLOGY

The purpose of this research is to present a methodology for the formulation of probability distributions of FM costs in a local NHS acute care trust building for the purposes a WLCC analysis. To perform this task, an original data sample of over 450 acute care NHS hospital Trusts’ FM costs was used as the basis of the study. The first stage was to remove samples from the set that the contained FM data on non-hospital sites. Non-hospital sites, such as primary care buildings, clinics etc were removed because they do not reflect the true FM costs of acute care hospitals resulting in distorted hypothesis testing of the distribution fitting later on in the research. Once this had been performed, eliciting a set of observed data that had on aggregate, an approximately equal mean floor area to that of a typical ward block building in a university teaching hospital, then reduced the sample further. Similarly, building characteristics such as gross heated volume were used to further reduce the sample. After consultation with practitioners, the final data sample was reduced to 52 sets to eliminate data sets that had missing or erroneous data. All final data sets exhibited similar characteristics to that of the main ward block building used as the basis for the study (i.e. similar heated volume, occupancy and floor area).

The data sets were then statistically analysed for distribution fitting, using two software applications, ExpertFit™ and BestFit™, resulting in a distribution for each FM cost centre. Testing the distribution against 28 continuous probability distributions and the Chi-square goodness-of-fit test validated each distribution. The two packages were used to compare results and identify any ambiguity in first ranked fits. It was found that both packages yielded almost identical results.

FITTING PROBABILITY DENSITY FUNCTIONS TO FM COSTS

Most approaches for economic risk analysis use subjective probabilities to describe the uncertainty of input variables when historical data may not be available (Ranasinghe and Russell 1993, Perry and Hayes 1985, Bjornsson 1977). However, when historical cost data is available, the collation of this data and the subsequent modelling of real-world scenario's can give rise to several problems when trying to create valid probability distributions. A simple heuristic technique for assessing the validity of a distribution is to plot a histogram of the data and visually inspect the variance, kurtosis and skewness of the data over the range (Law 1998). This can give a basic suggestion as to which distribution (or family of distributions) best represents the data, but there are several factors, which must also be addressed before selecting a possible distribution.

If probability distributions are to be used to create stochastic assumptions of WLCC inputs, then the way data is analysed and transformed into PDFs is of significant interest and importance. Weiler (1965) concluded that many errors in the outputs of simulation models could be traced back to assigning incorrect values to the parameters of a distribution, and indeed the selection of an appropriate distribution.

Where is it possible to collect data on whole life costs on FM cost centres of interest, such data can be used to specify a distribution based on one of the following approaches: a trace driven simulation, an empirical distribution, or a theoretical distribution function (Maio *et al.* 2000). If data is used to define an empirical distribution, the data is grouped to form a frequency histogram, and the resulting information is transferred to the simulation model. However, if the data set is used to fit a theoretical distribution using heuristics and goodness-of-fit techniques, it smoothes the irregularities that prevail and allows the possibility of sampling the extreme values of the distribution. This technique is regarded generally as the best method for performing simulations, and is used here for WLCC forecasts within the body of this research (Kirkham and Boussabaine 2000b).

PRE-DATA ANALYSIS: CLASS INTERVAL RULES

Class intervals, or "bins" are the ranges by which the data is grouped into on a histogram. The number and width of each class interval can have a significant bearing upon which distributions best fit the data being represented (when using the chi-square goodness-of-fit test). Some researchers (Montgomery and Ranger 1994) have suggested that the number of class intervals should fall in the region between five and twenty class intervals. They suggested that the square root rule should be used to calculate the number of observations. Simply, taking the square root of the number of observations in the data set derives the number of class intervals.

Sturges' Rule is reported on as another method of class interval selection (Maio *et al.* 2000). Sturges' Rule states that, for n observations, X_i to be summarized in a frequency distribution, then the number of class intervals for the distribution should be calculated by:

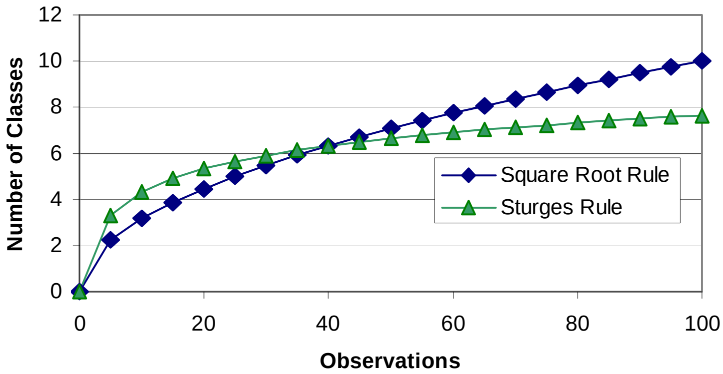
$$\mathrm{K} = \lfloor 1 + \log_2 n \rfloor \tag{1}$$

where K = number of bins and n = number of observations, and

$$M = \frac{[X_{max} - X_{min}]}{K} \tag{2}$$

where M = width of class intervals, X_{max} *and* X_{min} = maximum and minimum values of observations in the data set.

For the data presented in this paper, the selection of class interval rule is not critical as both methods yield similar results for distribution fitting and Chi-square evaluation. Figure 1 shows the convergence of both rules between thirty five to fifty five observations, which reflects the number of observations used in this research. However, for modelling using in excess of 70 observations, then careful consideration of which rule is most appropriate needs to be undertaken as the graph shows a clear divergence of class intervals after that point.



.

Figure 1: Comparison of Square root and Sturges' Rule for class interval calculation

GOODNESS OF FIT

Whichever method of calculating the number of class intervals in the distribution is used, the next stage is to fit a distribution to the data set. Although visual inspection can reveal which kind of distributions are most likely to represent the data, a statistical test should be performed to validate the choice of selected distribution.

The Chi-square test is a formal comparison of the relationship between the observed data set and the theoretical distribution fitted. The Chi-square test though is highly correlated to the class interval rule chosen and as such, the method used to calculate the number of class intervals effects upon the Chi-square test results, particularly so in data sets where more than seventy observations are used. This has led to the conclusion that the Chi-square test is weakened by its dependence on the class interval rule.

Notwithstanding, this test is widely used by construction researchers involved in fitting distributions to data sets. The chi-square statistic is defined as:

$$\chi^2 = \sum_{i=1}^{K} \frac{(N_i - E_i)^2}{E_i} \tag{3}$$

Where K = number of bins, N_i = the number of observed samples in the i^{th} bin and E_i = the expected number of samples in the i^{th} bin

Inspection of Figure 1 reveals that the class interval rule has no significant impact upon the accuracy of the distribution fitting procedure in this research. Sturges' rule and the square root rule converge at approximately 40 observations, and given the data set used in this study is 52, it can be concluded that either method will yield similar results, and not impact significantly upon the goodness of fit ranking procedure.

Table 1: Results of distribution fitting for each cost centre

Cost centre	Class interval rule	Observed data	First ranked distribution using the Chi-square Goodness of fit test
Water cost	Square root	52	Inverse Gaussian
Sewerage cost	Square root	52	Logistic
Clinical waste disposal cost	Square root	52	LogLogistic
Domestic waste disposal cost	Square root	52	LogLogistic
Meal provision cost	Square root	52	LogLogistic
Laundry and linen services cost	Square root	52	LogLogistic
Porterage cost	Square root	52	LogLogistic
Patient transport cost	Square root	52	LogLogistic
Non-stock items cost	Square root	52	Erlang
Cleaning cost	Square root	52	Pareto
Patient records cost	Square root	52	LogLogistic
Sterile services cost	Square root	52	Rayleigh
Postage costs	Square root	52	LogLogistic
Capital charges cost	Square root	52	Logistic
Security cost	Square root	52	LogLogistic
Telecommunications cost	Square root	52	LogLogistic

RESULTS

Table 1 shows the results of the fitting procedure. It can be observed that, when using the Chi-square goodness of fit test, the LogLogistic distribution was found to give first ranked fits for 62% of all distribution fits. However, in some cases where the LogLogistic distribution was not first ranked, the distribution was ranked sufficiently highly enough to consider its use in place of the first ranked distribution. It was therefore decided to assess whether the LogLogistic distribution could be used for the other cost centres where it was not the first ranked distribution (Water cost, sewage cost, non-stock items cost, cleaning cost and capital charges cost), through statistical justification. The purpose of this was to assess the validity of the hypothesis that the LogLogistic distribution is valid for all cost centres in FM services for acute care NHS hospitals.

For each cost centre that did not have a first ranked LogLogistic distribution, it was decided to compare both the first ranked distribution and a hypothesized LogLogistic distribution on a pair-wise basis. To do this, the parameters of the first ranked distribution were obtained from the results discussed in Table 1. Then, for the relevant cost centres, a LogLogistic distribution was also fitted and again, the parameters were elicited. Simulation software was then used to generate 100 random samples for each distribution within the defined parameters. The purpose of this was to test whether the hypothesized LogLogistic distribution differed significantly from the first ranked distribution, based on a randomly generated sample of 100 cost observations.

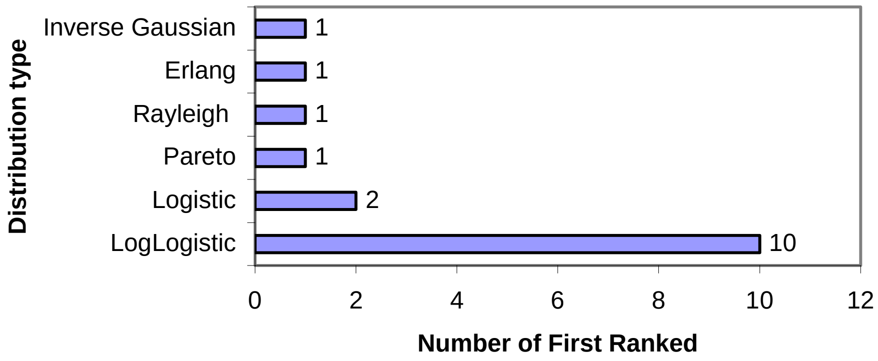


Figure 2: Frequency comparison of first ranked distributions for all cost centres

VALIDATION OF THE HYPOTHESIZED LOGLOGISTIC DISTRIBUTION

To test the validity of the hypothesized LogLogistic distribution, a simple hypothesis testing procedure was used. The distributions were compared using a two samples (assuming unequal variances) t-test. The t-test was used to examine a null and alternative hypothesis for each cost centre. If any of the two sets of distributions are similar, then the mean difference is expected to be around zero. But if the mean difference is much bigger than zero, then there will be a real difference between the distributions. Therefore, it is required that the null hypothesis of the mean difference being zero be tested. The following null hypothesis and alternative hypothesis were used for each cost centre.

1. $H_0:\mu_1=0$ the difference between the first ranked distribution and the hypothesized LogLogistic Distribution is small
2. $H_1:\mu_1 \neq 0$ the difference between the first ranked distribution and the hypothesized LogLogistic Distribution is large enough to suggest a real difference

Table 2 shows the statistics resulting from the t-test procedure. As a two-tailed test is used in this research, the $P(T<=t)$ two-tail value determines the acceptance region of he null hypothesis. For $P \geq 0.05$, the null hypothesis is accepted and for $P < 0.05$ the alternative hypothesis is accepted. It was found that for three out of six cost centres (water cost: $P(T<=t)$ two tail = 0.885, sterile services: $P(T<=t)$ two tail = 0.708 and sewage cost: $P(T<=t)$ two tail = 0.923) the LogLogistic distribution could be used in place of the first ranked distribution without any significant impact upon the accuracy of the results. The LogLogistic distribution could not be fitted to the capital charges cost centre because the distribution that best fitted this data is required to be of the unbounded continuous distribution type, not a non-negative distribution, such as the LogLogistic distribution. To reinforce the t-test results, the statistical differences between the first ranked distribution and the hypothesized Loglogistic distribution were calculated. To do this, a variety of statistical error measurements exist (Kirkham and Boussabaine 2000).

Recent forecasting research (Kirkham and Boussabaine 1999, Boussabaine *et al.* 1999) advocated the use of the Theils U statistic to test for differences between statistical models. Theils U calculates the difference by comparing changes in the observations of the first ranked distribution with changes in the hypothesized LogLogistic distribution. The U value is a coefficient value falling in the range between 0 and 2, where the difference is small as U tends to 0. The statistic is calculated using the following formula:

$$U = \frac{\sqrt{\dfrac{\sum_{t=1}^{n} e_t^2}{n}}}{\sqrt{\dfrac{\sum_{t=1}^{n} F_t^2}{n}} + \sqrt{\dfrac{\sum_{t=1}^{n} x_t^2}{n}}} \qquad (4)$$

where F = cost generated from hypothesized LogLogistic distribution, x = cost generated from first ranked distribution, n = number of observations and $e = x - F$

Table 2: Results of 2-sample assuming unequal variances t-test

Cost centre and distribution	Water cost centre		Sterile services cost centre		Sewage cost centre	
t-test descriptives	Inverse Gaussian	LogLogistic	Rayleigh	LogLogistic	Logistic	LogLogistic
Mean	83836.5073	83226.89968	357549.5966	368731.9895	82728.35819	83084.32622
Variance	884518111.1	906696376.7	45998796686	43436199931	689093296.9	685450348.2
Number of observations	100	100	100	100	100	100
Hypothesized mean difference		0		0		0
Degrees of freedom		198		198		198
t statistic		- 0.14403783		0.373921984		0.096013373
P(T<=t) one tail		0.442808543		0.354431227		0.46180353
t critical one tail		1.652585979		-		-
P(T<=t) two tail		0.885617086		0.708862454		0.923607059
t critical two tail		1.972016435		0.062786398		0.062786398

Cost centre and distribution	Non-stock items cost centre		Cleaning cost centre		Capital charges cost centre	
t-test descriptives	Erlang	LogLogistic	Pareto	LogLogistic	Logistic	LogLogistic
Mean	16906073.69	12296638.21	2980826.112	1344829.55		Invalid fit
Variance	3.0986E+14	9.16782E+13	4.5561E+13	1.95323E+11		
Number of observations	100	100	100	100		
Hypothesized mean difference		0		0		
Degrees of freedom		153		100		
t statistic		-2.30030013		-2.41855911		
P(T<=t) one tail		0.011391242		0.008695281		
t critical one tail		-		-		
P(T<=t) two tail		0.022782484		0.017390561		
t critical two tail		0.062809704		0.062864274		

Table 3: Error measurements statistics used in the validation process

Cost centre	*Error measurement statistics*					
	Mean error	*Mean absolute error*	*Sum of squared errors*	*Mean absolute percentage error*	*Theil's U statistic*	*McLaughlin's batting average*
Water cost	16.58148	16.58148	26947420	0.0011223	0.002814	399.7185
Sterile services	64.26773	64.26773	40481480	0.007425	0.001171	399.8829
Sewage cost	20.97798	20.97798	43131820	0.015076	0.000663	399.9337
Non-stock items	-291541	291541	8.33E+14	0.785195	1.0	250.8448
Cleaning cost	-289573.9	289573.9	8.22E+14	11.61338	2.0	<100
Capital charges	-	-	-	-	-	-

The table above gives the results of the tests. These statistical measurements supported the t-tests in that the water cost, sterile services and sewage cost all differed insignificantly from the hypothesized LogLogistic distribution, exhibiting U stats of 0.002, 0.001 and 0.00006 respectively, thus indicating almost identical fits.

DISCUSSION

Appendices 1, 2 and 3 show the final results of the fitting procedure post validation, providing the relative parameters and statistical descriptions. The two-stage validation process presented in this paper provided significant statistical justification for the use of the LogLogistic distribution in modelling the FM costs in acute care buildings in the NHS. After the hypothesis testing of non first-ranked LogLogistic distributions it was found that the distribution accounted for 81.25% of all first ranked distribution fits in all the FM cost centres. This was further supported by the analysis of both the variance of the skewness and kurtosis of each cost centre, which was relatively uniform, returning values of 0.515 and 8.503 respectively. Visual inspection of the distributions further supported this in that the costs were distributed principally around the lower end of the range of values, thus indicating a better fit for positively skewed distributions.

The results provide the analyst with a great deal of information about the costs of FM services when modelling the stochastic inputs into whole life cycle costing exercises. The use of empirical data in this research lends credibility to the assumption that the LogLogistic distribution is a suitable for representing FM costs. This is particularly useful in the analysis of whole life costs where empirical data to calculate the parameters of the distribution may not be available. Expert judgement can be used to form likely estimates of the parameters of the distribution, based upon the a priori assumption that the LogLogistic distribution best represents the FM costs.

CONCLUSION AND FUTURE WORK

The results of this research project support the overall hypothesis that the LogLogistic distribution is a powerful and accurate distribution for modelling all of the facilities management costs in acute care NHS buildings. Through a three-stage validation process, the accuracy of a hypothesized LogLogistic distribution was confirmed using historical facilities management cost data. These distributions can therefore be used as simulation inputs in whole life cycle costing exercises using Monte Carlo or Latin Hypercube simulation, for example, in forecasting the long term costs of FM services in acute care hospital buildings where historical data is unavailable to model assumptions. Generally, the triangular distribution is used in the absence of historical data but its weakness can have a significant impact upon the quality of simulation models. The knowledge gained from this research project provides evidence of the applicability a non-negative positively skewed distribution in this kind of cost modelling.

Using a case study, this paper has presented a methodology for developing PDF’s of FM costs for a specific building. Similar studies should be conducted on distribution fitting of FM costs within predefined ranges of dependent variables for those such as heated volume, number of occupants, gross floor area etc. For example, from the original data set, the total range of gross floor areas could be identified and then divided into distinct equal intervals. For each interval, distribution fitting could then be employed throughout all the cost centres to assess whether a) the type of distribution differs for each cost centre as the floor area increases and b) the type of distribution is homogeneous throughout all cost centres and all gross floor area ranges.

Identifying the type of PDF by ranges then provides practitioners with the ability to benchmark costs based on the characteristics of their own establishment, and offers the possibility to develop key performance indicators.

ACKNOWLEDGEMENTS

The authors would like to thank Mr Jonathan Millman and Mr Tim Litherland of NHS Estates, Mr Peter Jones of JPJ Energy Management and Mr Chris Joughin of The Royal Liverpool University Hospitals NHS Trust for their kind and valued co-operation in this research project.

REFERENCES

Audit Commission (2000) *The Acute Care Hospitals Portfolio.* London: The Audit Commission, HMSO.

Bjornsson, H.C. (1977) Risk Analysis of construction Cost Estimates. *Transactions*, American Association of Cost Engineers. 182-9

Boussabaine A H., Kirkham R. and Grew R. (1999) Modelling Total Energy Costs of Sports Centres. *Journal of Facilities Management.* **17**(12/13): 452-461.

BRE (1999) *Study on Whole life Costing for the Department of the Environment, Transport and the Regions*. Watford,United Kingdom: Building Research Establishment.

CBPP (1998) Factsheet on Whole Life Costing. *The Construction Best Practice Programme.*

Edwards, P.J. and Bowen, P.A. (1998) Risk and Risk Management in construction: a review and future directions for research. *Engineering Construction and Architectural Management.* **5**(4): 339-349.

Jovanovic, P. (1999) Application of sensitivity analysis in investment project evaluation under uncertainty and risk. *International Journal of project Management*. **17**(4): 217-222.

Kirkham R. J., Boussabaine, A. H. and Jones, J. P. (2000b) Modelling Energy Costs for Whole Life Costing in the NHS Estate. *Proceedings of the 2000 construction and building research conference of the Royal Institution of Chartered Surveyors (COBRA*), University of Greenwich.

Kirkham, R.J. and Boussabaine, A.H. (2000a) Developing a Framework for Whole Life Costing in the National Health Service Estate. *Proceedings of the 16th Annual Conference of the Association of Researchers in Construction Management*, Glasgow Caledonian University

Law, A.M. (1998) *Expert Fit User Manual*. Arizona, USA: Averill M. Law and Associates.

Maio, C., Schexnayder, C., Knutson, K. and Weber, S. (2000) Probability Distribution Functions for Construction Simulation. *Journal of Construction Engineering and Management*, ASCE. (July/August).

NHS (1999a) *The National Health Service Handbook 1999*. London: HMSO.

NHS Estates (1996) *Re-engineering the facilities management service, Health Facilities Note 16*. London: HMSO.

NHS Estates (1997). *The estate in the NHS, TFP Central Returns Data Analyses 1991/92-1995/96*. London: HMSO.

NHS Estates (1998*). A Business Approach to Facilities Management, Health Facilities Note 17*. London: HMSO.

NHS Estates (1999b) *Developing and Estate Strategy, Modernizing the NHS*. London: HMSO.

Perry, J.G. and Hayes, R.W. (1985) Risk and its management in construction projects. *Proceedings of the Institution of Civil Engineers.* UK, **78**(1): 499-521.

Ranasinghe, M. and Russell, A.D. (1993) Elicitation of subjective probabilities for economic risk analysis: an investigation. *Construction Management and economics*. **11**: 326-340

Weiler, H. (1965) The use of the incomplete beta functions for prior distributions in binomial sampling. *Technometrics*. **7**(3): 335-347.

Woodward, D. (1997) Life Cycle Costing – theory, information acquisition and application. *International Journal of Project Management*. **15**(6): 335-344

Appendix 1: Parameters and descriptive statistics for Logistic distribution

Cost Centre	Parameters			Descriptive Statistics						
	γ	β	α	Mean	Mode	Median	Standard Deviation	Variance	Skewness	Kurtosis
Water cost	-546206.08	632584.4074	36.5148	87159	85430	86378	31508	992743234	0.188[est]	3.6119[est]
Sewerage cost	-472819	558653.4	37.0882	86503	85022	85834	27393	7.5E+08	0.185[est]	3.6102[est]
Clinical waste cost	-245702	335194.4	16.84673	91446	87139	89496	36554	1.34E+09	0.409[est]	3.8147[est]
Domestic waste cost	-22600.4	58147.64	5.605916	38706	31924	35547	21215	4.5E+08	1.275[est]	5.9991[est]
Meal provision cost	-1085570.3	2089396.36	7.946355	1059265	938343	1003826	505644	2.56E+11	0.882[est]	4.7404[est]
Patient records cost	-1206764.9	1837627.637	8.8319	670194	584148	630863	395614	1.57E+11	0.790[est]	4.5098[est]
Laundry and linen cost	-145131	573785.9	5.477928	461359	391254	428655	215516	4.64E+10	1.308[est]	6.1206[est]
Porterage cost	-592866	1119743	8.910062	550419	498904	526877	238751	5.70E+10	0.783[est]	4.4928[est]
Patient transport cost	-1834664	2337802	24.57662	509517	495406	503138	173573	3.01E+10	0.279[est]	3.6778[est]
Sterile services cost	-1302790	1674678	13.08554	388085	352404	371888	237124	5.62E+10	0.528[est]	3.9857[est]
Postal services cost	-655655	778694.6	38.28339	123915	121978	123040	36984	1.37E+10	0.179[est]	3.6069[est]
Security cost	-28931.284	120182.4089	2.807842	120543	63245	91251	137088	1.88E+10	2.787[est]	14.387[est]
Telecommunication cost	-1019122	1421720	15.31985	412621	390526	402607	170955	2.92E+10	0.450[est]	3.8689[est]

Appendix 2: Parameters and descriptive statistics for Pareto distribution

Cost centre	Parameters		Descriptive Statistics						
	θ	A	Mean	Mode	Median	Standard Deviation	Variance	Skewness	Kurtosis
Cleaning cost	0.573087	233221	-	233221	781709				

Appendix 3: Parameters and descriptive statistics for Erlang distribution

Cost centre	Parameters			Descriptive Statistics						
	m	β	shift	Mean	Mode	Median	Standard Deviation	Variance	Skewness	Kurtosis
Non-stock items cost	0.573	233221	8165	13718542	8165	9511474	13710377	1.88E+14	2	9