NET CASH FLOW MODELS: TOWARDS THE BLACK BOX APPROACH

Ammar Kaka and Chris Fortune

School of the Built Environment, Heriot-Watt University, Riccarton Campus, Edinburgh, EH14 4AS, UK

Research in construction financial management has frequently centred on the development, usage and accuracy assessment of cash-flow forecasting tools and techniques. The conventional approach simulates S-curves of anticipated flows of cash in and out of the contracting organisation using step calculation. However, such approaches take no cognisance of the anticipated construction programme or detailed data contained in tender documentation and other soft issues such the effects of payment delays. This research addresses this matter by making an assessment of the feasibility of developing a black box approach that utilises simple multi-linear regression modelling. Data were simulated and examined in a pilot study using selected variables. The results of the research were inconclusive and it is suggested that further work in this area should focus non-linear regression models and ANNs

Keywords: cash-flow forecasting, simulation.

INTRODUCTION

Accurate cash flow forecasting is an essential activity during the bidding stage for all successful construction contracting organisations. It provides contractors with information such as:- the amount of capital required to perform a contract; the amount of interest needed to be paid to support any overdraft required; and the evaluation of different tendering strategies. Ideally, cash flow forecasts should be based on the anticipated construction programme and tendering documentation such as a bill of quantities. However, cash flow forecasting also needs to be simple and quickly produced considering the short time available during the tender period and the associated costs. A further difficulty is that contractors rarely prepare a comprehensive construction plan at the tendering stage, as they usually wait until winning the contract. Previous research has acknowledged the need for simple and fast techniques, and as a result simple ("first generation") cash flow forecasting models for contractors have been developed (see Mackay 1971; Ashley and Teicholz 1977; Kaka and Price 1991). These models are now incorporated in standard construction management text-books such as Harris and McCaffer 1999. These models tend to follow the same concept and mechanism. Standard S-curves which represent the running cumulative value of contracts (for example those developed by (Evans and Kaka 1998; Khosrowshahi 1991; Miskawi 1989; Bromilow and Henderson 1977; Hudson 1978) and are used to produce a running cumulative cost commitment curve by deducting the overall markup applied. These two curves are then converted (taking account of time delays and retention) into cash in and cash out profiles. The net of these curves gives the predicted cash flow for the contract.

Kaka, A P and Fortune, C (2002) Net cash flow models: towards the black box approach. *In:* Greenwood, D (Ed.), *18th Annual ARCOM Conference*, 2-4 September 2002, University of Northumbria. Association of Researchers in Construction Management, Vol. 1, 73-83.

Studies on the accuracy of cash flow models based on such ideal value curves have produced conflicting results. The feasibility of building ideal value curves for different project types is however, questionable. Kenley (1996) studied the variability of net cash flow profiles by collecting the cash in and cash out data from 26 projects. Comparison of the results found indicated that there were wide variations between the individual project profiles. Mackay's (1971) sensitivity analysis of net cash flow profiles to different value curves implied that either net cash flow curves conform to predictable patterns or that they were sensitive to the selection of payment delays. The sensitivity of the net cash flow profile to the selection of payment delays was studied by Kaka (1996) through a series of visits to 15 construction companies. These visits confirmed that time delays were usually controlled by contract conditions and their variability tended to be fairly limited. It was thus concluded that such basic models were not reliable and hence there was a strong justification for building a more flexible model, which could be adjusted to represent a wide range of variable profiles. As a result of the above, a "second generation" net cash flow model was developed (Kaka 1996). The objective of this model was to incorporate many more of the variables that feature in the actual cash flow mechanism. By examining contracts' conditions and the items involved in interim valuations, it was reported that many factors were lacking in the traditional or "first generation" models. Examples of these factors are those associated with tendering strategies, dates of interim valuations, subcontractors retention, estimating errors, etc. As a result, the revised model incorporated over 50 variables (compared to less than 10 in the traditional models). The model was tested and it was confirmed that many of the newly introduced variables were reasonably influential and hence their incorporation was necessary. The model was however still considered "simple" as it did not rely on the construction programme nor did it require data from bills of quantities. Moreover, the above study highlighted some of the limitations of adopting the simulation/calculation approach for cash flow forecasting. Both generations of models relied on a step by step simulation of the cash flow mechanism. To the authors' knowledge, no attempt has yet been made to forecast project net cash flow curves using an alternative "third generation" or "black box" approach. Such an approach is often useful when the mechanism or the process is either not fully understood or too complex to simulate (or calculate). This paper argues for the adoption of this approach and presents a simple multilinear regression model for project cash flow forecasting. The paper reports the results of a pilot study that is only a first stage of a research project aimed at developing and testing a third generation cash flow forecasting model based on this black box approach. The results reported in this paper are preliminary and hence should not be used to make a final judgement on the success or otherwise of the proposed approach.

WHY A BLACK BOX APPROACH?

The use of such black box approaches in construction management research is widespread. Whether the black box makes use of regression, artificial neural networks or genetic algorithms, construction management researchers have adopted these techniques to aid decisions on variety of management processes including cost forecasting (both capital and life cycle), duration estimation, tendering, etc. Indeed even in cash flow forecasting, regression and neural networks have been applied to problems related to the prediction of the shape of the S-curve (see Khosrowshahi 1991; Boussabaine and Kaka 1998). As mentioned above, these approaches are often useful when the mechanism or the process being modelled is either not fully understood or too complex to simulate. They are also applicable when detailed information/data are not available at the particular stage the model is to be used at. The study by Kaka (1996) pointed out some of the complexity associated in the cash flow mechanism. Figures 1 and 2 illustrate the cash flow mechanism identified and modelled by the above study. The simulation approach adopted to model the effect of all these variables on the cash flow output required the research to make some assumptions and simplifications and hence the model was liable to inaccuracy. The following points list the model parts that may be possible causes for inaccuracy:

1- The cost commitment S-curve assumes a mathematical formula. In real life cumulative monthly cost values conform to this profile but not precisely. Kenley and Wilson 1986 demonstrated this in previous work.
2- Splitting the contract cost curve into two parts where different mark-up rates are applied simulates unbalancing. Contractors in real life apply front-end loading at a Bill of Quantities (or schedules) levels of details.
3- Variations are simulated using an S-curve that is distributed over the running value of contract. In real life variations occur at discrete stages.
4- Time delays applied to individual cost headings are represented by percentages over 0 to 2 months. In real life these percentages are likely to alter during the life of the project and unless a detailed analysis of a contract's programme and tender documentation is performed it is not possible to assess these changes.

Traditional net cash flow models are in the main, based on the given contract conditions and whilst a "benchmark" cash flow forecast is needed for control purposes, contractors (and subcontractors) must predict as much as possible what is actually likely to happen in real life. For example, although forms of contracts specify the extent to which clients are allowed to delay payments to contractors, contractors must predict their cash needs based on what is likely to happen in practice. This necessitates the incorporation of client behaviour and a series of other soft issues. These issues together with project characteristics have significant impact on the levels of discrepancy found between interim measurement/payments and actual progress. The impact of disruption of work and claims must also be assessed and it is important that the cash flow model is capable of updating the forecasts required during the construction period. It is apparent from the above that the cash flow mechanism cannot be fully and accurately simulated unless the anticipated construction programme is incorporated. Even then, there are many inter-related "soft" factors that are difficult to assess and fully incorporate into a simulation model. The project participants and the project characteristics affect these factors. The existence of such a combination of factors makes the case for the adoption of a new "black box" approach to modelling.

THE PILOT STUDY

This paper proposes the adoption of multi-linear regression as an alternative technique to traditional cash flow forecasting models. The objective of this study is to assess the feasibility of this technique rather than develop a usable prototype. To this end, it was felt that a pilot study should be set up that used simple cash flow mechanisms and the data needed to develop the regression model would be generated using simulation. Once the approach is investigated and proven or otherwise to be feasible/accurate then it was felt that future work would need to be carried out to identify further influential variables so that a prototype model could then be based on real data.

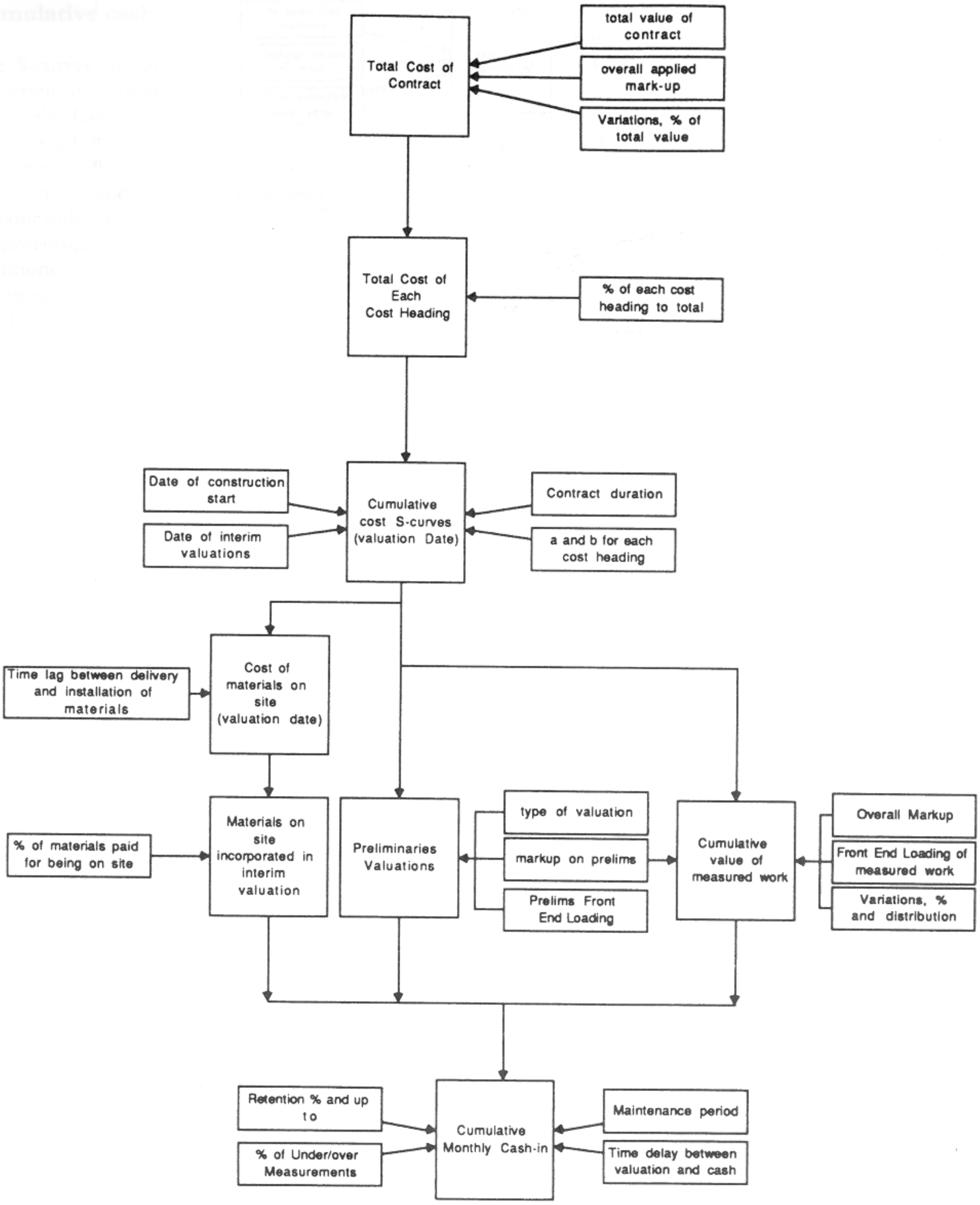

Figure 1: The mechanism of producing cumulative monthly cash in (source Kaka 1996)

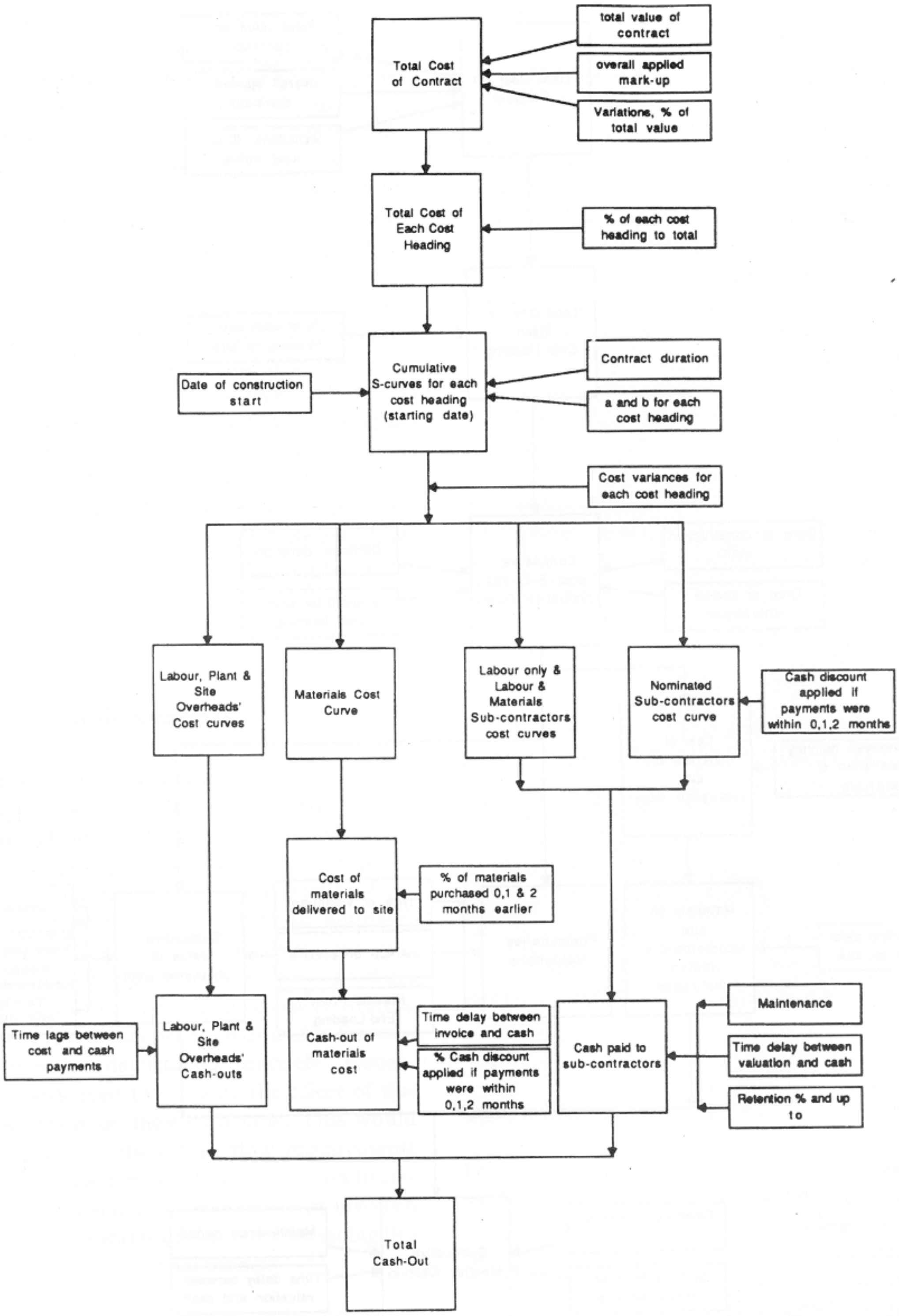

Figure 2: The mechanism of producing cumulative monthly cash out (source Kaka 1996)

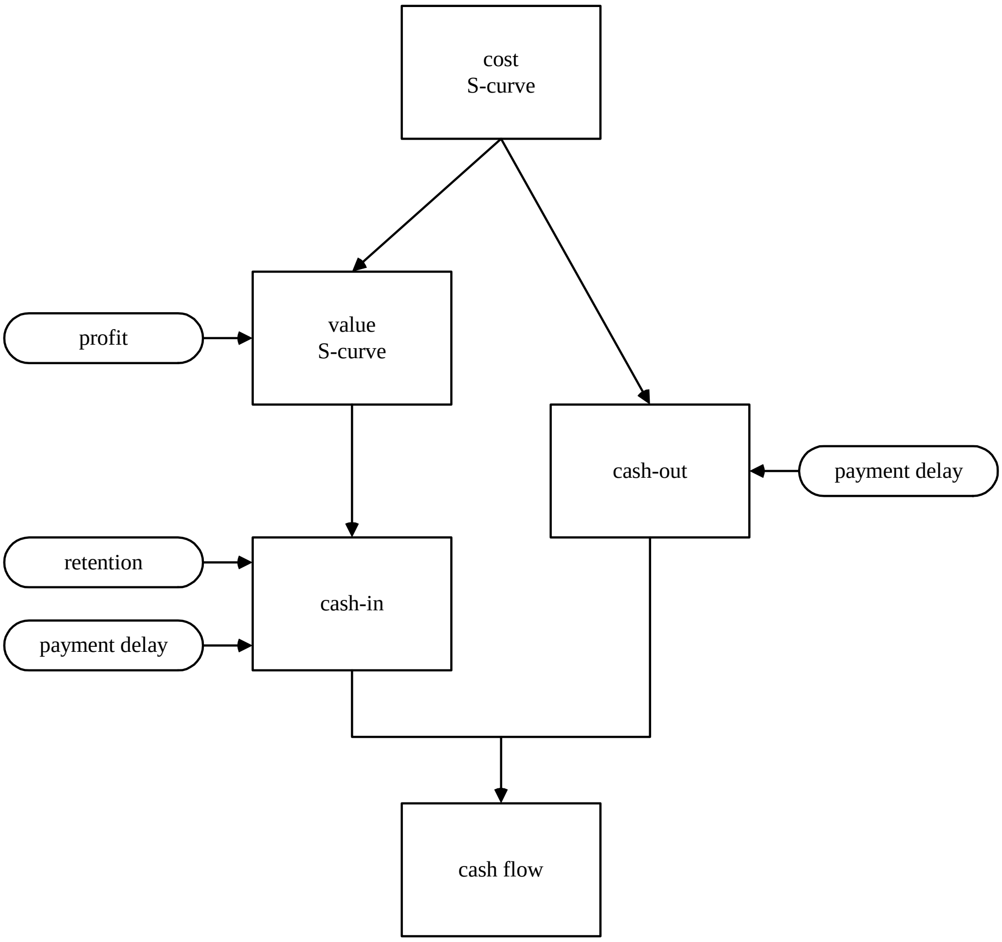

Figure 3: The traditional cash flow model

In order to test the feasibility of the proposed approach a traditional simulation model was developed using a spreadsheet. Figure 3 shows a simple cash flow mechanism with 11 data entry variables. The cost commitment S-cure is represented using the “logit transformation”. This is a technique that has been used in the past to fit an S-curve equation to cumulative cost or value flows of construction projects (Kenley and Wilson 1986; Kaka and Price 1993). This technique can be described as follows:

1- The linear equation is found by a logit transformation of both the independent and the dependent variables (in this case, time and cost respectively):

$$\text{Logit} = \text{Ln}\,\frac{Z}{1-Z} \tag{1}$$

Where Z is the variable to be transformed and Logit is the transformation.

2- Once the two variables are transformed, a linear equation can be fitted into the transformed data. This linear equation is expressed as follows:

$$Y = \alpha + \beta\, X$$

$$\text{Where } Y = \text{Ln } \frac{c}{100 - c} \text{ and } \quad X = \text{Ln } \frac{t}{100 - t} \qquad (2)$$

In order to transform data for a particular project, X and Y must be calculated for each value of t (time) and c (cost) respectively. Deriving the constants α and β is thus a simple linear regression of the transformed data.

3- Once α and β are derived, the fitted S-curve can be generated using the following equation:

$$c = \frac{100 \times F}{1 + F} \quad \text{Where } \quad F = \quad e^{\alpha}\left(\frac{t}{100 - t}\right)^{\beta} \qquad (3)$$

Values of α and β are considered to be entry variables for the purpose of this simulation model (see below for how these values were assigned). Multiplying the S-curve profile (represented in percentage format) then produces a cost flow curve by the total cost and duration of the project (2 data entry variables). Once a cost flow curve is generated, payment delays are applied to convert it to a cash-out curve. Assigning percentages of the costs that are delayed by 0 months, 1 months and 2 months (i.e. 3 data entry variables) simulated these. Applying the following formula (1 data entry variable) to the monthly cost values then produces the value flow curve:

$$\text{Value} = \quad \text{cost } (1 + \text{markup}) \qquad (4)$$

The value flow curve is then converted into the cash in curve by applying retention (rate and maintenance period being 2 data entry variables) and client payment delays (one data entry variable). Finally subtracting the cash out curve from the cash in curve derives the net cash flow curve.

Representing the model's output

The above model and the proposed black box model are developed to forecast the net cash flow of a construction project. The monthly cash balance for a given project assuming that a project cash account has commenced with a nil value ideally represents this output. Whilst this is an appropriate output for a traditional cash flow forecasting model, it is difficult to assign it as an output for a black box type of model. In a multi-linear regression model, outputs tend to be individual values rather than curves. Hence, a project with a duration of 12 months would essentially have 12 depended variables, whilst a 20 months project would have 20 variables. This is quite tedious and would distract from the use of this approach. A possible way forward would be to represent the net cash flow profile by a mathematical model/equation where constants would be used as substitute output. This is currently under investigation, but previous work indicates that net cash flow profiles vary significantly and a highly complicated model would be needed (Kaka and Price 1991).

To this end, it was decided that alternative dependent variables are to be identified to aid decisions related to projects' cash flow predictions. The following five factors are proposed:

Maximum negative cash flow (MNCF) (This would be useful to contractors to decide whether a project with given characterises and tendering strategies can be funded).

Net Present Worth (NPV) of the project's net cash flow (this is the financial out come of the project in terms of its worth at project commencement).

Compound amount (Comp) of project cash flow (this will give an indication as to the financial out come of the project at completion).

Equivalent annual cost (this transforms the project outcome into a uniform monthly value).

NPV/profit (this is a measure of the favourability of the cash flow).

Comp/Profit (this compares profit with actual outcome after interest is included).

All but one of the above dependent variables calls for the use of a discount or interest rate. As a result the total number of independent variables of the proposed model became 12.

Generating the training data

In order to develop/train the regression model, the above simulation model was run 100 times to generate the necessary data. Ranges were set for the data entry variables and the model was run by assigning uniform distributions to all but two (α and β) of these variables. The S-curve profiles were generated using a stochastic model developed by one of the authors in a previous study (Kaka 1999). In that study, the monthly cost commitment values of 118 completed building projects were collected from four building contractors. The objective was to develop a method in which the range (or envelope) of actual S-curves collected can be modelled and regenerated stochastically. Values of α and β were calculated for each project and hence following an application of equation (3) an S-curve was generated for each project using 10 time-intervals (uniform time-intervals to enable curves to be plotted together for analysis). Three time intervals were selected for the development of the stochastic models: at 30% of total duration, at 50% and at 70%. Results showed that the probability distributions for the 50% time-interval fitted closely to that of a normal distribution. The relationship between values at 50% and values at 30% in one hand and between values at 50% and values at 70% on the other hand were developed and plotted using linear regression.

As a result of the above study, S-curves generation starts with the values at 50% duration generated using the normal distribution. Then values at 30% and 70% would be calculated using the developed linear relationship plus the error of fit, which is generated using again, normal distributions. Therefore, for each run three values would be generated and consequently an S-curve function (logit transformation) is fitted to the three points and as a result values of α and β are calculated. Thus, for each run a set values of α and β and an S-curve are produced. This model was tested for accuracy in the previous study and was confirmed to be reliable (Kaka 1999).

The other data entry variables were assigned uniform distributions with the following ranges:

- Mark-up 1% to 10%
- Cost £800,000 to 4,000,000
- Duration 7 to 12
- Retention rate 2% to 10%
- Contractor's payment delay (0 months) 10% to 30%
- Contractor's payment delay (1 month) 60% to 70%
- Contractor's payment delay (2 month) 100% minus total of above
- Client's payment delay 0 to 2 months
- Interest 4% to 10%
- Maintenance period 4 to 6 months

The above ranges and distributions were not intended to be representative of the industry's practices, but were set to generate a series of data to determine whether the relationship between the independent and dependent variables can be accurately modelled using regression.

RESULTS AND DISCUSSION

For the purposes of this paper, all of the 12 independent variables were included in the regression model and there was no attempt to assess levels of accuracy as a result of omitting some (i.e. step-wise regression). Ninety sets of data (projects) were used to develop the models and the remaining ten were left for testing. A regression model was developed for each of the dependent variables and as a result 6 equations were obtained. Table 1 lists the regression coefficients together with accuracy measurement. The above table lists the R squared values of the 6 models. Whilst these values were relatively acceptable and encouraging, the subsequent application of these models to the ten test "projects" show significant discrepancy. The percentage error was used to measure the accuracy of the models' output. This is basically the difference between model output and actual output divided by the actual output.

Table 1: Regression and testing output

	MNCF	NPV	Comp	EAC	NPV/Profit	Comp/Profit
Mean % error	101%	41%	53%	70%	36%	38%
R Squared	0.828977	0.917138	0.91413	0.875941	0.733511	0.726404
Intercept	198568	-26480.7	-61526.3	-2450.69	0.135411	0.901685
α	-8079.21	-2226.06	-2831.41	-182.074	-0.00916	-0.01158
β	-97733.8	-6301.24	-8380.53	544.3603	-0.09972	-0.11451
Cost	-0.13289	0.04693	0.056143	0.00549	-5.3E-08	-6E-08
Mark-up	2215131	2223810	2625812	263049	5.250448	6.228909
Duration	43644.15	273.9565	2548.485	-1149.68	-0.00934	0.004799
Retention rate	-1077733	-226564	-288507	-35641.3	-1.49354	-1.84183
Client delay	-342609	-40687.4	-46903.5	-4913.84	-0.32539	-0.37613
Delay (0)	-612419	15964.31	37803	12128.35	0.546363	-0.21454
Delay (1)	110186.8	-17936.4	-38429.5	7472.7	1.389434	0.510499
Delay (2)	27683.75	74435.53	102555.6	20952.22	1.658893	1.041988
Interest	-62077.9	-164699	-119904	-17753.4	-0.84651	-0.45603
Maintenance	14414.55	-1428.4	-1980.94	-451.063	0.003105	0.007131

The mean percentage of error for each model is given in Table 1. This shows high values particularly to the prediction of the maximum negative cash flow value (i.e. 101%). Table 2 lists the percentage errors for each of the 10 testing projects.

The above results indicate that multi-linear regression (in the straightforward manner that has been applied) did not produce satisfactory results. Further attempts (in terms of selecting only some of the independent variables) may yield better results, but results lead to the conclusions that future attempts should explore other black box types of modelling such as ANN.

Table 2: Percentage error for test projects

Project	MNCF	NPV	Comp	EAC	NPV/Profit	Comp/Profit
1	1.173917	0.338178	0.37135	0.353867	0.016877	0.03886
2	0.159058	0.026448	0.012322	0.064266	0.156141	0.179379
3	2.323324	0.159191	0.158843	0.174245	0.449417	0.47192
4	2.422359	0.59252	0.561506	0.733098	0.150641	0.128863
5	0.122791	0.129021	0.070566	0.242908	0.138337	0.096051
6	0.611009	0.683484	0.705615	0.626105	0.168757	0.126085
7	0.178137	0.051799	1.285159	2.236022	1.907538	2.060764
8	2.418389	1.31629	1.360632	1.290423	0.156149	0.180617
9	0.162543	0.823632	0.777137	1.091222	0.392622	0.41291
10	0.524332	0.000164	0.00311	0.188492	0.152304	0.13798

CONCLUSIONS

This paper calls for the adoption of a third generation "black box" approach to cash flow forecasting. It has highlighted some of the reasons for proposing this approach as opposed to the traditional simulation/calculation approach. Some of these reasons have been found in the level of complexity associated with modelling real life cash flow mechanisms. Others are found in the nature of some of the factors (soft factors) affecting cash flow. The paper then attempted to test the feasibility of applying multi-linear regression to cash flow. A simple, "traditional" cash flow model was developed to generate the data needed to build a regression model. These data consist of 12 independent variables and 6 dependent variables. The dependent variables were identified to be important factors that the decision-maker would examine when having forecasted a project cash flow. The traditional model was run 100 times using assigned probability distribution and results obtained in previous studies. Ninety of these runs were used to develop/train the regression model and ten were used as testing projects. Whilst the R squared values obtained as a result of the regression model were acceptable, results of the tests yielded high levels of percentage errors.

The study concludes that multi-linear regression does not seem to be an applicable tool to develop contractors' cash flow mechanisms. Further studies should explore other types of "black box" techniques, such as non-linear regression or ANNs. If a modelling technique is found to be appropriate, then the next step would be to develop a prototype model that would be based on real data as opposed to simulated data. Such a model should also incorporate "soft" factors that are yet to be considered in cash flow research.

REFERENCES

Ashley D.B. and Teicholz P.M. (1977) Pre-estimate cash flow analysis. *Journal of the Construction Division, ASCE*, **103**, 369-79.

Boussabaine A. H. and Kaka A.P. (1998). A neural networks approach for cost flow forecasting. *Construction Management and Economics*, **16**, 471-479.

Bromilow, F.J. and Henderson, J.A. (1977). Procedures for reckoning the performance of building Contracts, 2nd edn. CSIRO, Division of Building Research, Highett, Australia.

Evans R.C. and Kaka A.P. (1998), Analysis of the accuracy of standard/average value curves using food retail building projects as case studies. *Engineering Construction and Architectural Management*, **5**(1), 58-67.

Harris, F.C. and McCaffer (1999). R. *Modern Construction Management*. 5th ed. Oxford: Blackwell Scientific

Hudson, K.W., (1978) DHSS expenditure forecasting method. *Chartered Surveyor - Building and Quantity Surveying Quarterly*, **5**,42-5.

Kaka A.P. and Price A. (1991) Net cash flow models: are they reliable? *Construction Management and Economics*. **9** 291-308.

Kaka A.P. and Price A. (1993) Modelling cost commitment curves for cash flow forecasting. *Construction Management and Economics.* **11**(4), 271-283.

Kaka A.P. (1996) Towards more flexible and accurate cash flow forecasting. *Construction Management and Economics*, **14**(1), 35-44.

Kaka A. P. (1999) The development of a benchmark model that uses historical data for monitoring the progress of current construction projects. *Engineering Construction and Architectural Management*, **6**(3), 256-267.

Khosrowshahi, F. (1991) *Simulation of Expenditure Patterns of Construction Projects. Construction Management and Economics*, **9**, 113-132.

Kenley R. (1996) *Construction project cash flow modelling*. PhD thesis, Department of Architecture and Building, University of Melbourne.

Kenley, R. and Wilson, O. (1986). A construction project cash flow model - an idiographic approach. *Construction Management and Economics*, **4**, 213-232.

Mackay, I. (1971) *To examine the feasibility of a computer programme for cash flow forecasting by contractors*. MSc Project in Construction Management at Loughborough University, UK.

Miskawi, Z. (1989). An S-Curve Equation for Project Control. *Construction Management and Economics*, **7**, 115-125.