A modified storey enclosure model

Cheung, Franco K.T.; Skitmore, Martin

Published in:
Construction Management and Economics

DOI:
10.1080/01446190500435093

QUT

QUTePrints
Open-access archive of QUT research literature

COVER SHEET

Franco, Cheung and Skitmore, R.M. (2006) A MODIFIED STOREY ENCLOSURE MODEL. *Construction Management and Economics* 24(3):pp. 391-406

A MODIFIED STOREY ENCLOSURE MODEL

Franco K T Cheung[1] and Martin Skitmore[2]

[1]Department of Building and Construction
City University of Hong Kong

[2]School of Urban Development
Queensland University of Technology

Corresponding Author:

Franco K T Cheung
Department of Building and Construction
City University of Hong Kong
Tat Chee Avenue
Kowloon
Hong Kong

16 August 2005 (version 3)

WORD COUNT (EXCLUDING FIGURES AND TABLES)

Title, Abstract and Keywords	279
Main text	4041
References	680

Total	5000

A MODIFIED STOREY ENCLOSURE MODEL

ABSTRACT

James' Storey Enclosure Method (JSEM), developed in 1954, is considered by many to be the most sophisticated single-rate method ever devised for early-design-stage tender price forecasts. However, the method is seldom used in practice partly because it has been superseded by multi-rate methods (such as the elemental method) and partly due to the arbitrary nature of the weightings prescribed for its use. This paper describes the further development of the approach, and in which empirical values of the weightings are derived by multivariate regression analysis.

A set of 50 completed Hong Kong private housing projects is used to demonstrate the use of the technique. This involves, firstly, the modification of the variables used in the original JSEM to incorporate the special characteristics of Hong Kong multi-storey residential buildings. This results in what is termed here as a Modified James' Storey Enclosure Model (MJSEM). Next, the optimal number of variables for inclusion in the model is identified by means of a dual stepwise cross validation regression procedure - resulting in a Regressed Modified Model for James' Storey Enclosure Method (RMJSEM). Also, using an amended version of MJSEM, the dual stepwise cross validation regression is used to produce a Regressed Modified Model for Amended Storey Enclosure Method (RMASEM).

The forecasting accuracy of RMJSEM and RMASEM is then compared with that of MJSEM together with the floor area and cube method to provide an indication of the improvement achieved. It is shown that the RMASEM provides significantly more consistent forecasts than the MJSEM and floor area models, leading to the conclusion that RMASEM may be the best model.

Keywords: Forecasts, cost model, regression, cross validation

INTRODUCTION

Many alternative approaches and new models have been developed for forecasting the tender price of individual projects. The majority of those reported focus on the uniqueness of a new model and the way in which it is different from other models (Raftery 1984; Newton 1990). Recent surveys of the United Kingdom (Fortune and Lees 1996; Fortune and Hinks 1998), South Africa (Bowen and Edwards 1998) and Nigeria (Akintoye *et al* 1992) however, show that conventional methods are still predominantly used in practice.

Generally, all conventional methods for forecasting building prices at the early design stage are single-rate methods. The first recorded of these was the cube method, which was invented approximately 200 years ago, with the more widely used floor area method being developed around 1920 (Skitmore *et al* 1990 p. xix). In response to criticisms that the floor area method failed to take into account such obvious features as building shape, height and number of storeys, James (1954) developed an alternative single-rate method, the storey enclosure method (JSEM), which uses a weighting scheme to quantify these characteristics into aggregated storey enclosure 'units' for pricing by a single appropriate storey enclosure unit rate. Although, as James admits, the weightings prescribed are arbitrary, his empirical analysis of the forecasting performance of his method demonstrated its superiority against both the cube and floor area alternatives – a subsequent recent reanalysis of his results showing this to be statistically significant at the 5% level (chi-square 5.99, 2df) (Skitmore 1991).

Despite these encouraging findings and its obvious simplicity in application, the JSEM today serves primarily as a textbook novelty, rather than a method that is used in practice. Survey results on the use of conventional cost forecasting models indicate that less than 2% of respondents in the UK (Fortune and Lees 1989) and 27% of the respondents in South Africa (Bowen and Edwards 1998) use the JSEM. Although there are no empirical data available concerning its lack of use in practice, several opinions have been voiced, including a lack of confidence in the arbitrarily prescribed weightings (Wilderness Group 1964; Ashworth 1999 p.251) and the lack of historical data available to estimate suitable unit rates (Wilderness Group 1964; Seeley 1996 pp.161-162). The subsequent development of multi-rate methods has also overshadowed interest in single-rate forecasting methods.

In the 50 year period since James' pioneering work, the availability of powerful personal computers and associated software has simplified the implementation of JSEM considerably. The use of such basic statistical techniques as multivariate regression analysis makes the task of estimating and validating weightings a relatively straightforward exercise and, as will be shown, renders the need for unit rates redundant. The potential of these techniques to further develop the JSEM is as yet unknown.

The results of research aimed at investigating this further by multivariate regression analysis are described in this paper and a set of 50 completed Hong Kong private housing projects is used to demonstrate the progress made. This involves, firstly, the modification of the variables used in the original JSEM to incorporate the special characteristics of Hong Kong multi-storey residential buildings. This results in what is termed as MJSEM. Next, the optimal number of variables for inclusion in the model is identified by means of a dual stepwise cross validation regression procedure - resulting in a regressed model termed as RMJSEM. Also, using an amended version of MJSEM, the dual stepwise cross validation regression is used to produce another regressed model termed as RMASEM. The forecasting accuracy of RMJSEM and RMASEM is then compared with that of MJSEM together with the floor area and cube method to provide an indication of the improvement achieved. It is shown that the RMASEM provides significantly more consistent forecasts than the MJSEM and floor area models, leading to the conclusion that RMASEM may be the best model.

JSEM FOR MULTISTOREY HOUSING IN HONG KONG

As noted in the Introduction, the JSEM takes into account various important aspects of design in building price forecasting, whilst leaving the type of structure and standard of finishes to be assessed in the price rate. Table 1 shows the various design aspects and the corresponding measurements used James' original model, represented by :

$$P = \left(\sum_{i=0}^{n} (2 + 0.15i) f_i + \sum_{i=0}^{n} p_i s_i + 2\sum_{j=0}^{m} f'_j + 2.5\sum_{j=0}^{m} p'_j s'_j + r \right) \cdot R \qquad (1)$$

where P is the forecasted price, R is the unit rate, f_i is the floor area at i storeys above ground, p_i is the perimeter of the external wall at i storeys above ground, s_i is the storey height at i storeys above ground, n is the total number of storeys above ground level, m is the total number of floors below ground level, f'_j is the floor area at j floors below ground level, p'_j is the perimeter of the external wall at j storeys below ground level, s'_j is the storey height at j storeys below ground level and r is the roof area.

In applying the JSEM to high-rise buildings, it is obvious that the higher the building, the more variables have to be created. If Equation (1) is used to estimate the price of a 30-storey building without a basement, then it will be necessary to measure the floor area, the perimeter and storey height thirty times (once for each level), the number of levels, and the roof area. 61 variables (e.g., p_0s_0, p_1s_1… and $p_{29}s_{29}$) therefore have to be calculated from 91 items of measurement (e.g., p_0, p_1… p_{29}, and s_0, s_1… s_{29}). However, including too many variables not only cause laborious works but also increases the possibility of counting irrelevant variables or of causing multicolinearity which results in numerically unstable models. Since the majority of Hong Kong multi-storey residential buildings comprise repeating floors and therefore the floor area is approximately the same at each level, this characteristic dramatically reduces the number of variables involved to just four: the total level, total elevation area (measured by multiplying the average perimeter by the overall building height), the average floor area and the roof area.

Equation (1) is adopted for modification. Let $\sum_{i=0}^{n} p_i s_i = np_{pt} s_{pt}$, where p_{pt} is the average perimeter of the superstructure and s_{pt}, is the average storey height of the superstructure floor. Let $\sum_{j=0}^{m} p'_j s'_j = mp_b s_b$, where p_b is the average perimeter of the basement and s_b, is the average storey height of the basement. Let, $\sum_{j=0}^{m} f'_j = mf_b$, where f_b is the average floor area per storey for floors at basement level and $f'_0 \approx f'_1 \approx \ldots \approx f'_m \approx f_b$ (the floor area for each level of the basement is more or less the same, and is approximately equal to f_b).

Typically, a building comprises a podium section and a tower section. Let $n = a + b$, where a is the number of storeys of the podium and b is the number of storeys of the tower. $f_0 \approx f_1 \approx \ldots \approx f_a \approx f_p$ (the floor area for each level of the podium is more or less the same, and is approximately equal to f_p), where f_p is the average storey area for floors at the podium level. $f_{a+1} \approx f_{a+2}, \ldots \approx f_b \approx f_t$ (the floor area for each level of the tower is more or less the same, and is approximately equal to f_t), where f_t is the average storey area for floors at tower level.

Then, $\sum_{i=0}^{n}(2+0.15i)f_i = \sum_{i=0}^{a+b}(2+0.15i)f_i = \sum_{i=0}^{a}(2+0.15i)f_p + \sum_{i=a+1}^{b}(2+0.15i)f_t$

$$= 2af_p + 0.15(0+1+\cdots+a)f_p + 2bf_t + 0.15\big[(a+1)+(a+2)+\cdots+(a+b)\big]f_t$$

$$= 2af_p + 0.15(0+1+\cdots+a)f_p + 2bf + 0.15abf_t + 0.15(1+2+\cdots+b)f_t$$

$$= 2af_p + 0.15\cdot\frac{a(a-1)}{2}f_p + 2bf_t + 0.15abf_t + 0.15\cdot\frac{b(b-1)}{2}f_t$$
$$= \left(2-\frac{0.15}{2}\right)af_p + \frac{0.15}{2}a^2 f_p + \left(2-\frac{0.15}{2}\right)bf_t + \frac{0.15}{2}b^2 f_t + 0.15abf_t$$

Equation (1) is therefore modified to MJSEM as follow:

$$\left.\begin{aligned} P &= \begin{bmatrix} \left(2-\frac{0.15}{2}\right)af_p + \frac{0.15}{2}a^2 f_p + \left(2-\frac{0.15}{2}\right)bf_t + \frac{0.15}{2}b^2 f_t + 0.15abf_t \\ + r + np_{pt}s_{pt} + 2mf_b + 2.5mp_b s_b \end{bmatrix}\cdot R \\ &= \left(2-\frac{0.15}{2}\right)af_pR + \frac{0.15}{2}a^2 f_pR + \left(2-\frac{0.15}{2}\right)bf_tR + \frac{0.15}{2}b^2 f_tR + 0.15abf_tR \\ &\quad + rR + np_{pt}s_{pt}R + 2mf_bR + 2.5mp_b s_b R \end{aligned}\right\} \quad (2)$$

IDENTIFICATION OF THE MODEL

Regression analysis

The arbitrary nature of James' formulation of the relationship between the building price and the variables in his method has often been criticized (e.g., Wilderness Group 1964; Ashworth 1999 p.251), even by James himself (James, 1954). It is likely that his model includes some irrelevant predicting variables, excludes some significant predicting variables such as the perimeter or the height of a building, or that the relationships exist between building prices and the predicting variables may be expressed in a better way than James' original formulation.

It is clearly impractical to examine this exhaustively, however. For a given set of data, there are always an unlimited number of possible explanatory models. Also, if a model is too simple, then its predictions may be unrealistic, whereas if a model is too complex, then although it may be more specific its predictions may be unreliable (Edwards 2001 p.129). Science convention is to make the model simple but not too simple – otherwise termed *the principle of parsimony* (Zellner et al. 2001).

One approach to this is by means of the variable selection algorithm used in the classical multiple linear regression analysis to determine the subset of variables and corresponding coefficients that provide the best prediction or forecast. Considering the James' original model, as amended for multi-storey Hong Kong housing in MJSEM, Equation (2) is easily converted to regression form as a model containing one dependent variable or response variable, P, with some independent or predictor variables, such as $nf_{pt}R$ and rR etc. Let all of the possible predictor variables be V_i, where $i = 1, 2, \ldots, k$, the building price model can be represented as:

$$P = \beta_0 + \beta_1 V_1 + \beta_2 V_2 + \cdots + \beta_k V_k = \beta_0 + \sum_{i=1}^{k} \beta_i V_i \qquad (3)$$

where β_0, β_is are constant coefficients and V_is are dependent variables.

Thus, the regressed model for MJSEM (RMJSEM or RMASEM) is the model comprising the (most important or most valid) subset of predictors from V_1 to V_k with the corresponding coefficients that gives the least mean square error (MSQ) in prediction.

Selection criterion for predictors

There are two approaches for selecting predictors based on errors of prediction – parametric and non parametric. The former approach refers to modern statistical inference that is based on the postulation of a parametric statistical model (Fisher 1922). The parametric models are arguably simpler than the non-parametric models because they are more informative, more amenable to statistical adequacy assessment, are often more parsimonious and are more likely to give rise to reliable and precise empirical evidence (Spanos 2001 p.186). Therefore, statistical adequacy can best be analysed in a parametric setting. However, the common assumption of normality that lies behind a parametric model may not be easily fulfilled in actual application.

With the exception of Skitmore and Patchell (1990), previous studies (e.g., Department of Environment 1971; Southwell, 1971; Tregenza 1972; Kouskoulas and Koehn 1974; Braby 1975; McCaffer 1975; Flanagan and Norman 1978; Karshenas 1984; Singh 1990; Khosrowshahi and Kaka 1996) using regression analysis all used parametric procedures although it is not always clear how well the data satisfy the statistical properties on which the procedure depends. Parametric procedures are arguably more suitable for routine problems with large sample size, but not for forecasting of building prices. Firstly, the number of projects available for analysis of building prices that are of a sufficiently similar nature and occur within a reasonable time span is usually rather small. This can easily cause the coefficients estimates to be biased. Secondly, the use of parametric techniques such as the least-squares method is known to be robust, even if the normality assumption is not fully satisfied.

However, the parametric estimates of the error rates may not be correspondingly robust (McLachlan 1987). Although variables can be transformed to fulfil the requirement of normality, the other assumptions such as unidimensionality may still cause violation.

A nonparametric alternative is to use the error of *ex ante* (out of sample) predictions to select variables and evaluate models. Of these, resampling methods make the least data demands. Three major resampling methods are available (Efron 1982): *cross validation*, in which one case is omitted in turn from the model derivation and the coefficients obtained from the analysis of the remaining cases applied to that case; the *jack-knife* method, in which one case is omitted in turn from the model derivation and the resulting coefficients are applied to the other cases; and the *bootstrap* method, in which the coefficients are used to generate simulated data from which a second set of coefficients is obtained. For predictive applications, the cross validation method has the most intuitive appeal as with non-time-series data of this nature each error value can be thought of as a real error that may arise in the practice of forecasting (Skitmore 1992). It also has the advantage of being markedly superior for small data sets (Goutte 1997).

The selection objective is set to minimize the average MSQ in fitting of the cross validated models. The accuracy of statistical inference when using cross validation is preserved by dividing a sample that contains n cases of data into n exploratory sub-samples (each containing n - 1 cases that are obtained from the original n-case sample by the omission of one case without repetition), each of which is used to select a statistical model using the least-squares approach, and n omitted cases, each of which is used to validate the selected model from an exploratory sub-sample that does not contain the omitted case. In this way, an average MSQ may be obtained from *n* models for each subset of candidate variables. The average MSQs from models of different subsets of candidates can then be compared, with the model with the smallest average MSQ being regarded as best subset model.

Selection strategy

To find the best subset by exhaustive means is a lengthy process and most often avoided by recourse to the three standard algorithmic selection procedures of forward inclusion, backward elimination and stepwise selection (a combination of forward and backward procedures) (Kleinbaum *et al* 1998 pp.392-399). Forward regression is applied by entering one candidate variable at a time. When no candidate that enters into the model can further reduce the average MSQ, the forward regression ends and the subset of variables that produced the minimal average MSQ is selected. Backward regression is applied by first entering all candidate variables and removing one variable at a time. When no candidate for removing from the model can further reduce the average MSQ, the backward regression ends and the subset of variables that produced the minimal average MSQ is selected. Stepwise regression involves applying forward and backward regression in turn until each ends. Stepwise regression ends when neither adding nor omitting a variable can further reduce the average MSQ. Clearly, stepwise regression is more thorough in its selection process. However, the results obtained can vary depending on whether the process starts with the forward phase or backward phase. One solution to this difficulty is to adopt a dual selection strategy by conducting the stepwise regression both ways and selecting the one with the best result (Fig. 1).

DEMONSTRATION WITH HONG KONG HOUSING DATA

A purpose-made programme was written using the mathematical software, Mathcad, to perform the resampling procedure and the selection algorithm. To identify the predictors for best subset models, the modelling process started off with the candidate variables that are used in MJSEM (See Table 2). The unit rate '*R*' in MJSEM (See Equation 2) was excluded, because the tender price is not measured on a unit area basis in regressed models. A new model, namely Regressed Modified Model for James' Storey Enclosure Method (RMJSEM), was developed by using the variables identified in MJSEM. The actual measurements of quantities (e.g. perimeter and storey height) for the variables in MJSEM (e.g. elevation area) were extracted to form the primary candidate variables for the regression analysis. A further model was developed, named the "Regressed Modified Model for Amended Storey Enclosure Method" (RMASEM), which uses another set of variables extracted from the candidate variables in MJSEM. RMASEM contains four types of candidates: (1) the primary variable (*n, m fpt, fb, spt, sb, ppt, pb, r*), (2) the second degree variable (*n2*), (3) the interaction term formed among primary variables (*nfpt, mfb, nspt, msb, nsptppt, msbpb*), and (4) the interaction term formed between primary variables and second degree variable (*n2fpt, n2spt, n2sptppt*).

When tender prices are used as the response for modelling, there is a risk of producing poorly performing models in terms of their percentage errors, i.e. the ratio of error (which is forecasted tender price minus the actual or lowest tender price) to the actual tender price. It is also found that the magnitude of error that is produced from forecasts of a wide range of tender prices varies significantly. As the performance of the forecasts are measured according to their percentage errors, the minimisation of total squared errors in the least-squares method is not necessarily an effective means of obtaining a good model unless tender prices in all of the cases in a model are fairly close to each other. To reduce the influence of a wide tender price range, the tender price per total floor area (*Y*) is adopted as the response. The tender price per total floor area is a sensible alternative because forecasters usually present building prices in unit prices, especially at the early budget stage, and the calculation of forecasted prices from the unit price model is straight forward. The unit price model can be directly compared with other conventional models despite their responses being different, because performance is measured on the basis of percentage errors.

Table 2 shows the response, the candidate variables and their corresponding equations for RMJSEM and RMASEM. The maximum model (that contains all candidate variables) for RMJSEM is:

$$\begin{aligned} Y' &= \beta_0' + \beta_1' afp + \beta_2' a2fp + \beta_3' bft + \beta_4' b2ft + \beta_5' abft + \beta_6' mfb + \\ &\quad \beta_7' nsptppt + \beta_8' msbpb + \beta_9' r \end{aligned} \tag{4}$$

and that for RMASEM is:

$$\begin{aligned} Y'' &= \beta_0'' + \beta_1'' n + \beta_2'' m + \beta_3'' n2 + \beta_4'' fpt + \beta_5'' fb + \beta_6'' spt + \beta_7'' sb + \beta_8'' ppt \\ &\quad + \beta_9'' pb + \beta_{10}'' nfpt + \beta_{11}'' n2fpt + \beta_{12}'' mfb + \beta_{13}'' nspt + \beta_{14}'' msb + \beta_{15}'' n2spt \\ &\quad + \beta_{16}'' nsptppt + \beta_{17}'' msbpb + \beta_{18}'' n2sptppt + \beta_{19}'' r \end{aligned} \tag{5}$$

where $Y^{'}$ and $Y^{''}$ are the forecasted price per total floor area for RMJSEM and RMASEM respectively, and β'_0, β'_1, … β'_9, and β''_0, β''_1, … β''_{19} are the corresponding coefficients determined by cross validation.

Tables 3-7 summarises the results of analyzing the Hong Kong Housing data (Appendix A) by cross validation regression in conjunction with the dual selection strategy. Table 3 shows the included candidates, excluded candidates, and selected predictors in RMJSEM and RMASEM with Tables 4 and 5 showing the step-by-step selection results of predictors for RMJSEM and RMASEM respectively, while Tables 6 and 7 shows the regression coefficients for each predictor, forecasts and MSQs for RMJSEM and RMASEM respectively.

Comparison of models

To appreciate the performance of the regressed models, RMJSEM and RMASEM, their forecast errors were compared with that of the three existing models, MJSEM, floor area and cube models. Using the data in Appendix A again to assess the performance of the models, tender price forecasts for MJSEM, floor area and cube models were calculated using Equations (2), (6) and (7) respectively. Equations (6) and (7) representing the floor area model (using total floor area of a building as the forecasting unit) and the cube model (using total volume of a building as the forecasting unit) respectively are shown as follows:

$$\hat{P}' = \left(a \cdot fp + b \cdot ft + m \cdot fb\right) \cdot R' \tag{6}$$

$$\hat{P}'' = \left(a \cdot fp \cdot sp + b \cdot ft \cdot st + m \cdot fb \cdot sb\right) \cdot R'' \qquad (7)$$

where, $\hat{P}'$ and $\hat{P}''$ are the forecasted prices for the floor area and cube models respectively and, R' and R'' are their corresponding unit rates obtained by cross validation.

Forecasting performance is judged by bias and consistency; bias is the arithmetic mean of percentage errors, and consistency is measured by the standard deviation of percentage errors. The higher the mean, the more biased is the model, and the higher the standard deviation, the less consistent is the model. Table 8 summarises the means and standard deviations of the percentage errors of all the models. As expected, the predictions from the cross validated models generally are not significantly biased. In terms of consistency counting from the best to the worst, the order of the models is RASEMH, cube, RJSEMH, floor area and. MJSEM. To judge whether the regressed models improve forecast over the conventional competing models, two groups of models were compared separately: Group A comprising RMASEM and the conventional models, and Group B comprising RMJSEM and the conventional models. Non-parametric tests were used as no transformations of the data could be found to the required normality assumption.

The Kruskal Wallis (K-W) test indicates that significant differences exist between in the consistency of models' forecasts within each group (Table 9). To identify the models responsible, pairwise examination of the models in each group was then made with the multiple two-sample Mann-Whitney *U*-test (Table 10). Unfortunately, however, performing multiple tests has a serious drawback; i.e., if each *U*-test has a 5% probability of erroneously rejecting the null hypothesis (H_0), then the probability of incorrectly rejecting at least one H_0 is much larger than 5%, and continues to increase with each additional test that is carried out (Kleinbaum *et al* 1998 pp. 443-447). Fisher's least significance difference (LSD) approach is used to correct exaggerated significance levels. The remedy is to rectify the significance level for each pairwise test to 0.083% (equivalent to 99.17% in the actual testing procedure).

Fig. 2 provides a graphical presentation of the LSD comparisons for the models in Group A and B. The models are ranked, in ascending order of sample variances, RMASEM, the cube model, the floor area model and MJSEM model respectively in Group A; with the cube model, RMJSEM, the floor area model and MJSEM respectively in Group B. In particular, RMASEM in Group A was very consistent (15.95%) for an early stage estimator. In this group, RMASEM and the cube model have the same potency; the cube and floor area models have the same potency; the floor area model and the JSEM have the same potency; both RMASEM and the cube model differs from MJSEM; RMASEM differs from the floor area model. Therefore, the more consistent set of models comprises the two comparable models: RMASEM and the cube model. In Group B, the cube model, RMJSEM and floor area model have the same potency; RMJSEM, floor area model and MJSEM have the same potency; the cube model differs from the MJSEM. Therefore, the more consistent set of models in Group B comprises the cube model, RMJSEM and floor area model. The result of two group comparisons reveals that MJSEM is the worst model. The better regressed model, RMASEM, that is the most consistent and significantly better than floor area model and MJSEM, is arguably the best model whereas the cube model is a competitive alternative which is less consistent than RMASEM but the difference is statistically insignificant. For application purposes, the best subset

RMASEM predictors (determined by the leave-one-out cross validation) are trained again by the least square error method using all project data to produce the following equation:

$$Y_0 = -6369.897 - 0.127 \cdot fpt + 0.608 \cdot fb + 3858.605 \cdot spt - 129.686 \cdot sb - 2.996 \cdot pb \quad (8)$$

where Y_0 is the forecasted price per total floor area.

CONCLUSIONS

In revisiting James' Storey Enclosure Method (JSEM), it has been shown how his model can firstly be presented in mathematical form and then how this form simplifies to a modified JSEM for typical Hong Kong private sector multi-storey housing projects (MJSEM) and then easily converts for regression analysis for empirical reestimation of the weighting coefficients involved. Using a dual stepwise cross validation regression, a best set of predictor variables is identified for a set of completed housing projects for both a Regressed Model for MJSEM (RMJSEM) and Regressed Modified Model for Amended Storey Enclosure Method (RMASEM). Statistical hypothesis tests on the forecasting performance of MJSEM, RMJSEM and RMASEM, together with the equivalent floor area and cube models lead to the conclusion that RMASEM is arguably the best model, with MJSEM and floor area model being least consistent.

The documented reluctancy of practitioners to use more sophisticated early stage approaches together with their increased adoption of powerful computers and software suggests that further research and development of the JSEM is likely to be well received in practice. Once developed, both experienced and inexperienced forecasters should be easily able to use the models with the aid of a simple spreadsheet. With a little training and practice, it should even be possible for practitioners to follow the methodology described in this paper to develop and validate their own cost models without the need for specialist statistical advice.

A major benefit of the approach is in the use of cross validation regression to simultaneously build and evaluate a range of potential models through the simulation of their likely performance on projects outside the sample base. This has considerable intuitive appeal as it produces forecasts in a similar way to forecasters, i.e. extracting the most relevant information from a pool of historical projects to make a prediction, while at the same time maximising the objectivity of the process in achieving, for example, variable parsimony. Of course, this does not preclude the user's professional judgement, which can still be applied in the form of a subjective final adjustment. It does, however, offer a step towards maximizing the use of whatever patterns exist in the data prior to such an adjustment.

REFERENCES

Akintoye, A. S., Ajewole, O. and Olomolaiye, P. O. (1992) Construction cost information management in Nigeria. *Construction Management and Economics* **10**: 107-116.

Ashworth, A. (1999) *Cost studies of buildings*. Harlow, England: Longman.

Bowen, P. A. and Edwards, P. J. (1998) Building Cost Planning and Cost Information Management in South Africa. *International Journal of Procurement*

(June): 16-25.

Braby, R. H. (1975) *Costs of high-rise buildings*. Building Economist. **14:** 84-6.

Department of Environment (DOE) (1971) *Local Authority Offices: Areas and Costs*. DOE, London.

Edwards, A. W. F. (2001) Occam's bonus. In A. Zellner, H. A. Keuzenkamp and M. McAleer (eds.). *Simplicity, inference and modelling: keeping it sophisticatedly simple*. Cambridge, Cambridge University Press**:** 128-132.

Efron, B. (1982) *The jackknife, the bootstrap, and other resampling plans*. Philadelphia, Pa.:, Society for Industrial & Applied Math..

Fisher, R. A. (1922) On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society* **A 222**: 309-368.

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

Fortune, C. and Hinks, J. (1998) Strategic building project price forecasting models in use - paradigm shift postponed. *Journal of Financial Management of Property and Construction* **3**(1): 3-26.

Fortune, C. and Lees, M. (1989) *An investigation into Methods of Early Cost Advice for Clients*, Salford College of Technology.

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

Goutte, C. (1997) Note on free lunches and cross-validation. *Neural Computation*

9(6): 1246-9.

James, W. (1954) A New Approach to Single Price Rate Approximate Estimating. *RICS Journal.* **XXXIII (XI)**(May): 810-24.

Karshenas, S. (1984) Predesign Cost Estimating Method for Multistory Buildings. *Journal of Construction Engineering and Management, ASCE* **110**(1): 79-86.

Khosrowshahi, F. and Kaka, A. P. (1996) Estimation of Project Total Cost and Duration for Housing Projects in the U.K. *Building and Environment* **31**(4): 375-383.

Kleinbaum, D. G., L. L. Kupper, et al. (1998) *Applied regression analysis and other multivariable methods*. Boston, Mass: PWS-Kent.

Kouskoulas, V. and Koehn, E. (1974) Predesign cost estimating function for buildings. *Journal of Construction Division, ASCE*: 589-604.

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

McLachlan, G. J. (1987) *Advances in multivariate statistical analysis*. A. K. Gupta. Dordrecht, Holland, D. Reidel.

Newton, S. (1990) An agenda for cost modelling research. *Construction Management and Economics*.

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

Seeley, I. H. (1996) *Building economics: appraisal and control of building design* *cost and efficiency.* Houndmills, Basingstoke, Hampshire: Macmillan,.

Singh, S. (1989) Computer Model for Cost Estimation of Structures in High Rise Commercial Buildings. *Proceedings of the 25th Annual Conference - Associated Schools of Construction*.

Spanos, A. (2001) Parametric versus non-parametric inference: statistical models and simplicity. In A. Zellner, H. A. Keuzenkamp and M. McAleer (eds.). *Simplicity, inference and modelling: keeping it sophisticatedly simple*. Cambridge, Cambridge University Press**:** 181-206.

Skitmore, M. (1992) Parameter prediction for cash flow forecasting models. *Construction Management and Economics* **10**: 397-413.

Skitmore, M., Stradling, S., Tuohy, A. and Makwezalamba, H. (1990) *The Accuracy of Construction Price Forecasts: A Study of Quantity Surveyors' Performance in Early Stage Estimating*. The University of Salford.

Skitmore, R. M. (1991) *Early Stage Construction Price Forecasting - A Review of Performance*. The Royal Institution of Chartered Surveyors.

Skitmore, R. M. and Patchell, B. R. T. (1990) Development in contract price forecasting and bidding techniques. In P. S. Brandon (ed.) *Quantity Surveying Techniques: New Directions*. Blackwell Scientific**:** 75-120.

Southwell, J. (1971) *Building Cost Forecasting*. The Royal Institution of Chartered Surveyors.

Tregenza, T. (1972) Association between building height and cost. *Architects' Journal* **156**(44): 1031-2.

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

Zellner, A., Keuzenkamp, H. A. and McAleer, M. (2001) *Simplicity, inference and modelling: keeping it sophisticatedly simple*. Cambridge, Cambridge University Press.

APPENDIX A: COST DATA

The data below were provided by an established surveying practice in Hong Kong and comprise the cost analyses of average grade private Hong Kong housing projects from a ten-year period commencing with the 3rd quarter of 1988.

Case	No. of floor for podium	No. of floor for tower	No. of floor for basement	Avg. area per floor for podium	Avg. area per floor for tower	Avg. area per floor for basement	Avg. storey height of podium	Avg. storey height of tower	Avg. storey height of basement	Avg. perimeter on plan for podium and tower	Avg. perimeter on plan for basement	Roof area	Original tender price*	Date of returned tender	Tender price index**	Adjusted tender price*** (tp x TPI / 1660)
	(a)	*(b)*	*(m)*	*(fp)*	*(ft)*	*(fb)*	*(sp)*	*(st)*	*(sb)*	*(ppt)*	*(pb)*	*(r)*	(tp)		*(TPI)*	(TP)
1	1	39	0	4960	2920	0	4.50	2.80	0.00	1110	0	4960	6.2E+08	Mar-97	1575	6.5E+08
2	1	41	0	7960	3590	0	4.20	2.80	0.00	1370	0	7960	9.3E+08	Apr-97	1660	9.3E+08
3	3	21	0	3070	1030	0	3.50	2.89	0.00	319	0	3070	1.7E+08	Jan-97	1575	1.8E+08
4	7	32	0	1010	433	0	3.14	3.50	0.00	144	0	1010	1.1E+08	Dec-96	1520	1.2E+08
5	1	33	0	2300	5440	0	4.20	2.70	0.00	1420	0	2300	5.2E+08	Mar-96	1400	6.2E+08
6	4	52	2	24900	2170	14000	3.50	2.75	3.40	564	2440	24900	1.0E+09	Nov-96	1520	1.1E+09
7	2	14	0	350	150	0	4.00	2.80	0.00	78	0	350	1.2E+07	Jan-95	1280	1.6E+07
8	3	13	0	696	306	0	3.33	2.75	0.00	123	0	696	2.0E+07	Feb-94	1100	3.0E+07
9	3	44	0	3110	765	0	3.67	2.80	0.00	272	0	3110	1.2E+08	Jan-94	1100	1.9E+08
10	4	28	1	15000	2290	10100	3.23	2.75	3.37	990	1450	15000	5.4E+08	Oct-94	1220	7.4E+08
11	0	37	1	0	5320	14600	0.00	2.75	3.37	1410	2120	5320	8.1E+08	May-93	1045	1.3E+09
12	0	33	1	0	3690	10300	0.00	2.75	3.63	983	1580	3690	3.4E+08	Nov-92	1030	5.4E+08
13	6	24	0	2390	833	0	3.17	3.15	0.00	235	0	2390	1.3E+08	Aug-93	1075	2.0E+08
14	4	29	0	346	131	0	3.38	2.90	0.00	66	0	346	1.8E+07	Mar-93	1025	3.0E+07
15	6	20	0	489	244	0	3.17	3.15	0.00	85	0	489	2.3E+07	Feb-93	1025	3.7E+07
16	3	38	0	10800	4340	0	3.50	2.70	0.00	1690	0	10800	4.0E+08	Oct-90	1080	6.2E+08
17	4	11	0	338	131	0	3.25	2.70	0.00	62	0	338	7.2E+06	Aug-90	1045	1.1E+07
18	3	38	0	5300	3300	0	3.33	2.70	0.00	826	0	5300	3.3E+08	Feb-90	1020	5.3E+08
19	2	16	0	910	314	0	3.45	2.63	0.00	127	0	910	1.5E+07	Nov-89	1000	2.5E+07
20	0	21	1	0	1360	2960	0.00	3.10	3.14	469	544	1360	9.9E+07	May-89	960	1.7E+08
21	3	35	1	2270	1690	865	3.63	3.10	3.22	557	168	2270	2.0E+08	Oct-89	1000	3.4E+08
22	0	34	0	0	1350	0	0.00	2.75	0.00	484	0	1350	1.2E+08	Dec-88	925	2.2E+08
23	2	33	2	2310	1540	2350	3.25	2.75	3.11	461	551	2310	1.1E+08	Oct-88	925	1.9E+08
24	2	36	0	12500	4510	0	3.50	2.70	0.00	1330	0	12500	3.8E+08	May-88	820	7.7E+08
25	4	20	0	328	123	0	1.98	2.80	0.00	61	0	328	9.6E+06	Oct-88	925	1.7E+07
26	2	15	1	7350	5100	4570	3.40	2.67	3.63	1580	799	7350	1.8E+08	Jun-88	820	3.6E+08
27	0	16	0	0	1290	0	0.00	2.67	0.00	381	0	1290	4.6E+07	Oct-88	925	8.3E+07
28	4	37	0	3810	2040	0	3.30	2.70	0.00	745	0	3810	1.3E+08	Nov-87	740	2.9E+08
29	1	23	1	4340	4500	787	4.00	2.75	3.31	1340	138	4340	1.9E+08	May-88	820	3.9E+08
30	2	18	1	2130	1560	2130	3.60	3.00	3.47	469	445	2130	8.3E+07	Dec-87	740	1.9E+08
31	0	32	1	0	3760	2430	0.00	2.75	3.90	1090	303	3760	1.8E+08	Nov-87	740	4.1E+08
32	0	31	0	0	1050	0	0.00	2.70	0.00	373	0	1050	4.4E+07	Oct-87	740	9.9E+07
33	3	36	0	8700	1350	0	3.40	2.70	0.00	678	0	8700	1.7E+08	Nov-87	740	3.7E+08
34	3	30	0	5420	2100	0	3.27	2.70	0.00	830	0	5420	1.2E+08	Nov-87	740	2.8E+08
35	3	36	0	5500	2260	0	3.47	2.70	0.00	706	0	5500	1.9E+08	Apr-88	820	3.9E+08
36	2	29	4	932	488	1480	3.60	2.70	3.25	175	303	932	4.0E+07	Nov-87	740	9.0E+07
37	0	37	0	0	3350	0	0.00	2.53	0.00	708	0	3350	3.3E+08	Jun-96	1430	3.8E+08
38	0	39	0	0	4340	0	0.00	2.70	0.00	1270	0	4340	3.3E+08	Nov-90	1080	5.1E+08
39	0	37	0	0	2320	0	0.00	2.70	0.00	941	0	2320	1.7E+08	Oct-90	1080	2.6E+08
40	0	37	0	0	4560	0	0.00	2.70	0.00	1690	0	4560	3.2E+08	Mar-90	1020	5.1E+08
41	0	37	0	0	3470	0	0.00	2.70	0.00	1270	0	3470	2.7E+08	Aug-90	1045	4.2E+08
42	0	36	0	0	2430	0	0.00	2.70	0.00	904	0	2430	2.3E+08	Aug-90	1045	3.7E+08
43	0	37	0	0	3510	0	0.00	2.70	0.00	1540	0	3510	2.9E+08	Sep-90	1045	4.5E+08
44	0	35	0	0	603	0	0.00	2.65	0.00	196	0	603	4.2E+07	Oct-88	925	7.5E+07
45	0	37	0	0	1910	0	0.00	2.65	0.00	611	0	1910	1.8E+08	Mar-96	1400	2.2E+08
46	1	24	0	1600	1100	0	0.00	2.67	0.00	371	0	1600	3.9E+07	Apr-88	820	7.9E+07
47	0	26	0	0	956	0	0.00	2.75	0.00	303	0	956	5.8E+07	Nov-87	740	1.3E+08
48	0	19	1	0	754	610	0.00	2.95	3.37	253	118	754	3.7E+07	May-89	960	6.4E+07
49	0	14	0	0	1450	0	0.00	2.90	0.00	530	0	1450	4.9E+07	Jul-90	1045	7.8E+07
50	0	39	0	0	3300	0	0.00	2.75	0.00	1030	0	3300	2.4E+08	Apr-90	1030	3.9E+08

Remarks:

* - Original tender price excludes the prices for all foundation, building services, external works, preliminaries and contingencies.

** - Index figures refer to the quarterly publication - Tender Price Indices and Cost Trends produced by Levett and Bailey Chatered Quantity Surveyors Ltd.

*** - Tender price is rebased to the price level in the 2nd quarter of 1997 (Tender Price Index = 1660)

CAPTIONS

Figures

1. Algorithm for Dual Stepwise Selection
2. Tests of Homogeneity of Variances Using Kruskal Wallis Tests and Mann-Whitney U-Tests

Tables

1. Measurements in JSEM to represent the design aspects that affect building prices
2. Candidates, Responses and their Equations for RMJSEM and RMASEM Included Candidates, Excluded Candidates and Selected Predictors for RMJSEM and RMASEM
3. Step-by-Step Selection Results of Predictors for RMJSEM
4. Step-by-Step Selection Results of Predictors for RMASEM
5. Coefficients, Forecasts and MSQs Determined by Leave-One-Out Method for RMJSEM
6. Coefficients, Forecasts and MSQs Determined by Leave-One-Out Method for RMASEM
7. Summary of Means and Standard Deviations of Percentage Errors
8. Kruskal Wallis (K-W) tests between Models
9. Summary of Means and Standard Deviations of Percentage Errors
10. Two-sample Mann-Whitney U-tests between Models

Table 1: Measurements in JSEM to represent the design aspects that affect building prices

Design aspects that affect building prices	**Measurements**
Shape of building	External wall Area
Floor area	Area of each floor
Vertical positioning of the floor area in a building	Greater multiplier (weighting) assigned to the floor area of a suspended floor positioned higher in a building
Storey heights of building	Proportion of floor and roof areas to the external wall area
Overall Building heights	Ratio of roof area to external wall area
Extra cost of sinking usable floor area below ground level	Constant multiplier assigned to floor area and external wall area below ground level

Table 2: Candidates, Responses and their Equations for RMJSEM and RMASEM

Variable	Equation	Notation
RMJSEM		
Candidates		
Total floor area for podium	$a \cdot fp$	*afp*
Storey number for podium · Total floor area for podium	$a^2 \cdot fp$	*a2fp*
Total floor area for tower	$b \cdot ft$	*bft*
Storey number for tower · Total floor area for tower	$b^2 \cdot ft$	*b2ft*
Storey number for podium · Total floor area for tower	$a \cdot b \cdot ft$	*abft*
Total floor area for basement	$m \cdot fb$	*mfb*
Elevation area	$(a \cdot sp + b \cdot st) \cdot ppt$	*nsptppt*
Basement wall area	$m \cdot sb \cdot pb$	*msbpb*
Roof area	r	*r*
Response		
Adjusted tender price per total floor area	$P \div (a \cdot fp + b \cdot ft + m \cdot fb)$	*Y*
RMASEM		
Candidates		
Storey number for superstructure	$a + b$	*n*
Storey number for basement	m	*m*
Square of storey number for superstructure	$(a + b)^2$	*n2*
Average area per storey for superstructure	$(a \cdot fp + b \cdot ft) \div (a + b)$	*fpt*
Average area per storey for basement	fb	*fb*
Average storey height of superstructure	$(a \cdot sp + b \cdot st) \div (a + b)$	*spt*
Average storey height of basement	sb	*sb*
Average perimeter on plan for superstructure	ppt	*ppt*
Average perimeter on plan for basement	pb	*pb*
Total floor area for superstructure	$(a \cdot fp + b \cdot ft)$	*nfpt*
Storey number for superstructure · Total floor area for superstructure	$(a + b) \cdot (a \cdot fp + b \cdot ft)$	*n2fpt*
Total floor area for basement	$m \cdot fb$	*mfb*
Height of building above ground	$(a \cdot sp + b \cdot st)$	*nspt*
Depth of basement	$m \cdot sb$	*msb*
Storey number for superstructure · Height of building above ground	$(a + b) \cdot (a \cdot sp + b \cdot st)$	*n2spt*
Elevation area	$(a \cdot sp + b \cdot st) \cdot ppt$	*nsptppt*
Basement wall area	$m \cdot sb \cdot pb$	*msbpb*
Storey number for superstructure · Elevation area	$(a + b) \cdot (a \cdot sp + b \cdot st) \cdot ppt$	*n2sptppt*
Roof area	r	*r*
Response		
Adjusted tender price per total floor area	$P \div (a \cdot fp + b \cdot ft + m \cdot fb)$	*Y*

Table 3: Included Candidates, Excluded Candidates and Selected Predictors for RMJSEM and RMASEM

RMJSEM		***RMASEM***	
afp	o	*n*	o
a2fp	o	*m*	o
bft	**o**	*n2*	o
b2ft	o	*fpt*	**o**
abft	o	*fb*	**o**
mfb	o	*spt*	**o**
nsptppt	o	*sb*	**o**
msbpb	x	*ppt*	o
r	**o**	*pb*	**o**
		nfpt	x
		n2fpt	x
		mfb	o
		nspt	o
		msb	x
		n2spt	x
		nsptppt	o
		msbpb	x
		n2sptppt	x
		r	x

Legend:
o - Candidate
x - Excluded Candidate
o - Selected Predictor
NA - Not applicable

Table 4: Step-by-Step Selection Results of Predictors for RMJSEM

Forward Stepwise

Step	Variables entered	Variables deleted	Average MSQ
1	*bft*		9.72E+05
2	*r*		9.59E+05
3		*(No entry or deletion, end regression)*	
Final model:	***bft, r***		**9.59E+05**

Backward Stepwise

Step	Variables entered	Variables deleted	Average MSQ
1	*afp, a2fp, bft, b2ft, abft, mfb, nsptppt, r*		1.69E+06
2		*mfb*	1.33E+06
3		*b2ft*	1.15E+06
4		*nsptppt*	1.08E+06
5		*a2fp*	1.06E+06
6		*abft*	9.95E+05
7		*afp*	9.59E+05
8		*(No deletion or entry, end regression)*	
Final model:	***bft, r***		**9.59E+05**

Table 5: Step-by-Step Selection Results of Predictors for RMASEM

Forward Stepwise

Step	Variables entered	Variables deleted	Average MSQ
1	*spt*		5.96E+05
2	*fb*		5.62E+05
3	*pb*		5.19E+05
4	*fpt*		4.96E+05
5	*sb*		4.92E+05
6	*(No entry or deletion, end regression)*		
Final model:	***spt, fb, pb, fpt, sb***		**4.92E+05**

Backward Stepwise

Step	Variables entered	Variables deleted	Average MSQ
1	*n, m, n2, fpt, fb, spt, sb, ppt, pb, mfb, nspt, nsptppt*		5.26E+06
2		*m*	6.74E+05
3		*nspt*	6.11E+05
4		*ppt*	5.67E+05
5		*n*	5.41E+05
6		*mfb*	5.20E+05
7		*nsptppt*	5.02E+05
8		*n2*	4.92E+05
9	*(No deletion or entry, end regression)*		
Final model:	***spt, fb, pb, fpt, sb***		**4.92E+05**

Table 6: Coefficients, Forecasts and MSQs Determined by Leave-One-Out Method for RMJSEM

Case	RMJSEM (β_0 + $\beta_1 \cdot bft$ + $\beta_2 \cdot r$)				
	β_0	β_1	β_2	Forecasted Y	MSQ
1	4530	-0.008	0.057	3,935	2.41E+06
2	4567	-0.008	0.052	3,784	4.77E+06
3	4484	-0.007	0.055	4,504	1.53E+06
4	4466	-0.007	0.058	4,430	2.09E+06
5	4533	-0.007	0.057	3,353	8.83E+02
6	4496	-0.008	0.087	5,744	1.27E+06
7	4475	-0.007	0.058	4,480	1.27E+06
8	4509	-0.007	0.057	4,520	2.25E+05
9	4536	-0.007	0.056	4,466	8.91E+03
10	4531	-0.007	0.045	4,771	5.08E+05
11	4625	-0.010	0.069	2,997	9.64E+06
12	4533	-0.007	0.057	3,845	5.93E+04
13	4474	-0.007	0.056	4,471	1.99E+06
14	4470	-0.007	0.058	4,464	1.52E+06
15	4522	-0.007	0.057	4,514	4.95E+04
16	4509	-0.007	0.061	4,024	7.96E+05
17	4558	-0.007	0.056	4,566	2.35E+05
18	4532	-0.007	0.056	3,926	3.08E+04
19	4583	-0.008	0.056	4,595	9.98E+05
20	4490	-0.007	0.058	4,368	1.12E+06
21	4511	-0.007	0.058	4,213	6.41E+05
22	4516	-0.007	0.058	4,261	2.60E+05
23	4568	-0.007	0.055	4,321	1.35E+06
24	4535	-0.007	0.056	4,048	6.96E+03
25	4530	-0.007	0.056	4,531	3.55E+03
26	4541	-0.007	0.060	4,416	4.88E+05
27	4552	-0.007	0.056	4,472	1.99E+05
28	4551	-0.007	0.056	4,219	9.78E+05
29	4536	-0.007	0.056	4,037	1.91E+05
30	4495	-0.007	0.057	4,418	9.56E+05
31	4534	-0.007	0.055	3,890	2.68E+05
32	4586	-0.008	0.054	4,399	1.85E+06
33	4526	-0.007	0.054	4,648	1.10E+05
34	4551	-0.007	0.059	4,403	8.02E+05
35	4537	-0.007	0.057	4,259	9.87E+04
36	4552	-0.007	0.056	4,499	1.61E+05
37	4534	-0.007	0.054	3,856	6.24E+05
38	4523	-0.007	0.054	3,600	3.37E+05
39	4550	-0.007	0.054	4,067	1.01E+06
40	4523	-0.007	0.054	3,614	3.22E+05
41	4533	-0.007	0.055	3,820	2.76E+05
42	4530	-0.007	0.057	4,028	3.60E+04
43	4532	-0.007	0.055	3,802	9.83E+04
44	4572	-0.007	0.055	4,447	7.69E+05
45	4558	-0.007	0.054	4,153	1.11E+06
46	4601	-0.008	0.055	4,486	2.78E+06
47	4499	-0.007	0.058	4,377	6.62E+05
48	4542	-0.007	0.056	4,479	3.55E+04
49	4560	-0.007	0.056	4,490	4.01E+05
50	4532	-0.007	0.054	3,822	6.27E+05
Average:					**9.59E+05**

Table 7: Coefficients, Forecasts and MSQs Determined by Leave-One-Out Method for RMASEM

Case	RMASEM (β_0 + $\beta_1 \cdot spt$ + $\beta_2 \cdot fb$ + $\beta_3 \cdot pb$ + $\beta_4 \cdot fpt$ + $\beta_5 \cdot sb$)							
	β_0	β_1	β_2	β_3	β_4	β_5	Forecasted Y	MSQ
1	-6090	3757	0.617	-3.045	-0.142	-114.9	4,166	1.74E+06
2	-5808	3666	0.632	-3.119	-0.165	-105.6	3,969	4.00E+06
3	-6091	3748	0.603	-2.975	-0.125	-118.9	4,868	7.60E+05
4	-8495	4625	0.633	-3.110	-0.116	-167.1	7,330	2.12E+06
5	-6419	3873	0.606	-2.985	-0.121	-131.6	3,564	3.29E+04
6	-6386	3863	0.514	-2.298	-0.126	-150.0	5,053	1.91E+05
7	-6281	3813	0.600	-2.955	-0.117	-123.0	4,946	4.37E+05
8	-6403	3861	0.605	-2.978	-0.120	-126.6	4,590	1.64E+05
9	-6370	3861	0.612	-3.015	-0.128	-130.7	4,538	2.76E+04
10	-6369	3857	0.581	-2.836	-0.126	-131.7	5,292	3.67E+04
11	-6309	3836	0.454	-2.157	-0.127	-121.2	5,212	7.90E+05
12	-6208	3804	0.664	-3.268	-0.133	-122.0	4,995	8.22E+05
13	-6162	3782	0.608	-2.995	-0.128	-124.7	5,620	6.93E+04
14	-6258	3803	0.599	-2.949	-0.116	-122.3	4,974	5.24E+05
15	-7189	4168	0.629	-3.097	-0.134	-148.9	5,918	1.40E+06
16	-6508	3901	0.598	-2.948	-0.110	-136.2	3,724	3.51E+05
17	-6301	3846	0.618	-3.048	-0.137	-132.8	4,622	2.92E+05
18	-6373	3859	0.610	-3.005	-0.126	-129.8	3,790	1.51E+03
19	-6145	3791	0.617	-3.042	-0.138	-130.8	4,118	2.73E+05
20	-6251	3815	0.620	-3.069	-0.126	-134.9	5,148	7.65E+04
21	-6459	3890	0.611	-3.025	-0.126	-124.1	5,164	2.27E+04
22	-6544	3908	0.606	-2.984	-0.120	-125.6	4,042	5.32E+05
23	-6197	3799	0.543	-2.588	-0.130	-119.2	3,631	2.22E+05
24	-6249	3822	0.623	-3.071	-0.145	-122.6	3,518	3.76E+05
25	-6855	4009	0.599	-2.947	-0.106	-129.1	3,806	6.15E+05
26	-6368	3863	0.628	-3.110	-0.133	-135.9	3,451	7.07E+04
27	-6496	3898	0.608	-2.998	-0.123	-128.7	3,753	7.48E+04
28	-6291	3837	0.610	-3.006	-0.127	-134.9	4,014	6.16E+05
29	-6369	3860	0.610	-2.991	-0.129	-134.5	3,488	1.24E+04
30	-6114	3767	0.662	-3.302	-0.127	-151.2	4,623	5.97E+05
31	-6217	3801	0.676	-3.492	-0.124	-91.8	3,994	3.85E+05
32	-5997	3744	0.617	-3.046	-0.140	-132.3	3,964	8.56E+05
33	-6525	3902	0.608	-2.997	-0.124	-123.1	3,983	9.94E+05
34	-6329	3847	0.609	-3.003	-0.126	-132.6	3,955	2.01E+05
35	-6371	3859	0.610	-3.007	-0.127	-129.5	3,958	1.81E+02
36	-6601	3934	0.619	-3.043	-0.118	-146.3	3,707	1.53E+05
37	-6433	3880	0.611	-3.012	-0.127	-129.2	2,957	1.18E+04
38	-6380	3859	0.601	-2.964	-0.116	-133.9	3,536	2.67E+05
39	-6199	3804	0.609	-3.003	-0.127	-133.2	3,777	5.09E+05
40	-6392	3862	0.602	-2.965	-0.116	-133.6	3,509	2.14E+05
41	-6341	3848	0.607	-2.991	-0.123	-131.8	3,622	1.07E+05
42	-6486	3895	0.611	-3.012	-0.127	-126.7	3,722	2.46E+05
43	-6361	3855	0.609	-3.001	-0.125	-130.3	3,608	1.43E+04
44	-6237	3817	0.612	-3.020	-0.132	-129.6	3,797	5.16E+04
45	-6142	3785	0.611	-3.012	-0.131	-131.0	3,638	2.90E+05
46	-5902	3705	0.614	-3.031	-0.139	-128.7	3,441	3.88E+05
47	-6694	3952	0.600	-2.954	-0.111	-124.3	4,067	1.26E+06
48	-6355	3855	0.616	-3.069	-0.130	-117.2	4,541	6.25E+04
49	-6510	3918	0.615	-3.031	-0.129	-137.7	4,664	6.52E+05
50	-6383	3864	0.603	-2.972	-0.117	-136.5	3,856	6.81E+05
Average:								**4.92E+05**

Table 8: Summary of Means and Standard Deviations of Percentage Errors

	Mean % error (m)	*p*-value for *t*-test (H_0: m=0)	Standard Deviation of % error	Predictors for Regressed Models
RMASEM	2.66%	0.24	15.95%	*fpt, fb, spt, sb,pb*
CUBE	1.47%	0.60	19.59%	-
RMJSEM	4.84%	0.14	22.64%	*bft, r*
FLOOR AREA	1.31%	0.69	23.53%	-
MJSEM	-2.73%	0.51	29.04%	-

Table 9: Kruskal Wallis (K-W) tests between Models

	Kruskal Wallis (K-W) tests (at 95%* significance level)			
Group	*Chi-square*	*Df*	*p*-value	H_0: No difference in absolute deviation (reject if p < 0.05)
A	23.879	3	0.000	**Reject H_0**
B	12.918	3	0.005	**Reject H_0**

Table 10: Two-sample Mann-Whitney U-tests between Models

	Mann-Whitney *U*-test (at 99.17%* significance level)		
Pair	*Z*	*p*-value	H_0: No difference in absolute deviation (reject if p < 0.0083)
<u>Comparisons for Group A</u>			
MJSEM and RMASEM	**-4.3707**	**0.0000**	**Reject H_0**
Floor Area and RASEM	**-3.0126**	**0.0026**	**Reject H_0**
Cube and RMASEM	-1.7441	0.0811	Accept H_0
<u>Comparisons for Group B</u>			
MJSEM and RMJSEM	-2.4818	0.0131	Accept H_0
Floor Area and RMJSEM	-1.1651	0.2440	Accept H_0
Cube and RMJSEM	-0.4481	0.6541	Accept H_0
<u>Common Comparisons for Group A and B</u>			
MJSEM and Floor Area	-1.8544	0.0637	Accept H_0
Floor Area and Cube	-1.6821	0.0926	Accept H_0
Cube and MJSEM	**-3.0609**	**0.0022**	**Reject H_0**

Remark: * – 99.17% = (1 - 0.05/6 (no. of pairs)) x 100%

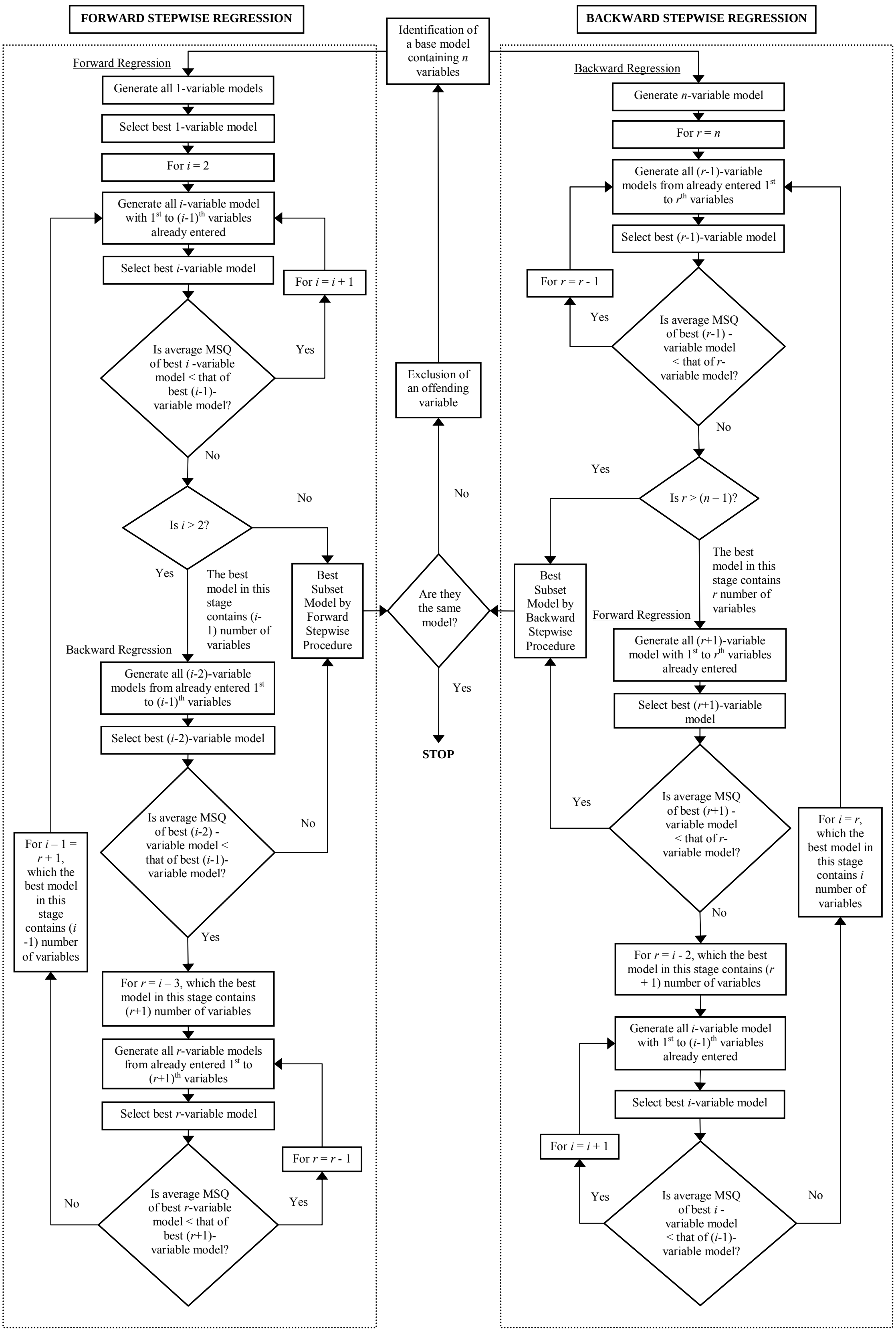
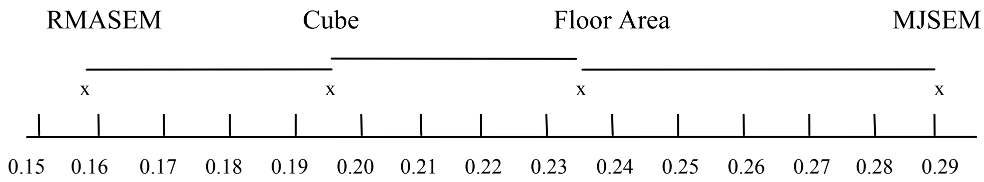


Fig. 1: Algorithm for Dual Stepwise Selection

<u>LSD Comparison of Sample Variances (by U-tests) for Group A</u>
<u>(RMASEM and the conventional models)</u>



<u>LSD Comparison of Sample Variances (by U-tests) for Group B</u>
<u>(RMJSEM and the conventional models)</u>

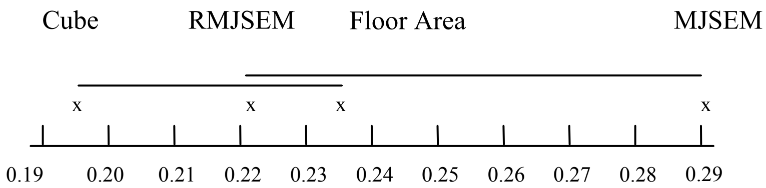


Fig. 2: Tests of Homogeneity of Variances Using Kruskal Wallis Tests and Mann-Whitney U-Tests