DATA POOLING FOR EARLY-STAGE PRICE FORECASTS

Daisy K.L. Yeung[1] and Martin Skitmore [2]

[1]*Department of Building and Construction, City University of Hong Kong.H.K.*
[2]*School of Urban Development, Queensland University of Technology, Brisbane, Queensland 4001, Australia*

Many clients or building owners rely heavily on the early stage construction cost forecasts, provided by the Quantity Surveyor (Q.S.), for their investment decisions and advance financial arrangements. During the preliminary stage, the information concerning the new project is very scarce. It is very common in practice for Q.S. to use historical building cost data on which to base the forecast of the new project – typically basing the forecast on the known price of the most similar project to the new one. However, this approach not always generates the most accurate results. This paper develops an idea for early stage forecasting by using out-of-sample mean square errors to measure the forecasting accuracy. A method is presented for finding the best pooling arrangement of the available data source according to the characteristics of the new project. By making the best use of the available data pool, it can maximize the quality of the early stage cost forecast with limited project information.

Keywords: data group pooling, forecasting accuracy, homogeneity, mean square error, price family.

INTRODUCTION

Many clients or building owners rely heavily on the early stage construction cost forecasts (particularly in the feasibility study stage), provided by the Quantity Surveyor (Q.S), for their investment decisions and advance financial arrangement. During the preliminary design stage, it is very common in practice for the Q.S. to use historical building cost data on which to base the forecast of the new project - typically basing the forecast on the known price of a similar project to the new one. Earlier researchers have pointed out the issues concerning forecasting accuracy with this and the possibilities of using a group of similar projects instead of a single project. For clarity, we shall refer to a new project as the *target-project* or just *target*; a single similar project as the *base-project*; a group of similar projects as the *base-group* of projects; and the process of selecting the projects that comprise the base-group is called *data pooling*.

FAMILY OF PRICES

Statistician Beeston (1975) states that a potentially more powerful method than the base-project approach is to use the mean price of suitably composed base-group on the rationale that the target project price is one of a "family of prices" for that project, and for which the prices of those in the base-group are a proxy. In this case,

[1] bcdaisyy@cityu.edu.hk
2 rm.skitmore@qut.edu.au

Yeung, D K L and Skitmore, M (2005) Data pooling for early-stage price forecasts. *In:* Khosrowshahi, F (Ed.), *21st Annual ARCOM Conference*, 7-9 September 2005, SOAS, University of London. Association of Researchers in Construction Management, Vol. 1, 269-76.

the mean of the base-group provides the best estimate of the prices (observations) in the base-group by minimising the standard error involved (and maximum likelihood if the sample is normal). This being the case, it is easy to why this may be extended to a (future) observation outside the sample.

HOMOGENEITY OF COST DATA

The major difficulty with data pooling is what Flanagan (1980) has termed the "homogeneity" problem, in that the bigger the base-group the less similar are the projects in the base-group. In statistical parlance this means that, although the increased sample size reduces the standard error of the mean at the same time, the sample becomes less representative of the population from which the target project is conjectured to belong. In other words, making the best use of available data involves a trade off between the need to restrict the base-group used to that most relevant to the target project while at the same time maximizing the base-group size.

In 2001, Skitmore offered an approach to solving this in the risk analysis context by empirically examining the effects of all possible pooling combinations on forecasting errors with a view to selecting the data pooling arrangement that best minimises the spread of errors.

PROPOSED POOLING OF AVAILABLE HISTORICAL PRICE DATA

The research described in this paper aims to propose a method to define the best pooling arrangement of available data to suit the characteristic of the target project and by making the best use of available data to maximize the accuracy and performance of the cost forecast. Making the best use of available data involves: (1) a trade off between the need to restrict the base-group used to the most relevant to the target project, while at the same time (2) maximizing the size of the base-group involved.

There are different ways to organise the data source for forecasting the target project. In Skitmore's pooling study (2001), it was proposed to group the cost data into 5 groups for reviewing, they are (1) Construction floor area, (2) Contract sum, (3) Nature of works, (4) Project type, and (5) Number of bidders. In this study, we restricted ourselves to just three of these: (1) Construction floor area (CFA), (2) Building type, and (3) Client type. Based on these groups, source data are pooled into sub-groups. They are pooled into sub-groups according to : (1)construction floor area (CFA), (2) type of client, (3) type of building and (4) combination of (1), (2) and (3).

Table 1: Data pooling table.

Main Group	CFA	Type of Client	Type of Building	Combination of CFA, Client Type and Building Type	All	Total Nos. of sub-groups
Nos. of Sub-group	6	2	30	462	1	500

Table 2**: Sub-groups according to CFA**

S	S & M	S,M & L (ALL)
M	S & L	
L	M & L	

Where S – Small CFA sub-group, M – Medium CFA sub-group and, L – Large CFA sub-group

Table 3**: Sub-groups according to types of client**

Private Sector Client (Pr)	Public Sector Client (Pu)	Private and Public Sector Client (All)

Table 4**: Sub-groups according to types of building**

R	R,Sh	Sh, P	R, Sh, C	R,P, Sc	R, Sh, C,P
Sh	R,C	Sh, Sc	R, Sh, P	Sh, C, P	R, Sh, C, Sc
C	R,P	C,P	R, Sh, Sc	Sh, C, Sc	R, Sh, P, Sc
P	R, Sc	C,Sc	R, C, P	Sh, P, Sc	R, C, P, Sc
Sc	Sh, C	P,Sc	R, C, Sc	C, P, Sc	Sh, Sc, C, P

R – Residential, Sh – School, C – Commercial center, P – Carpark, Sc – Social community center

Table 5: **Sub-groups for all combination**

S, Pr	S,M, Pr	S,Pu	S,M,Pu	S,R	S,M,R	S,Sh	S,M,Sh	S,C	S,M,C
M,Pr	S,L,Pr	M,Pu	S,L,Pu	M,R	S,L,R	M,Sh	S,L,Sh	M,C	S,L,C
L,Pr	M,L,Pr	L,Pu	M,L,Pu	L,R	M,L,R	L,Sh	M,L,Sh	L,C	M,L,C
S,P	S,M,P	S,Sc	S,M,Sc	S,Pr,R,P,Sc	S,M,Pr,R,P,Sc	S,Pr,R,Sh,C,P	S,M,Pr,R,Sh,C,P	S,Pr,R,Sh,C,Sc	S,M,Pr,R,Sh,C,Sc
M,P	S,L,P	M,Sc	S,L,Sc	M,Pr,R,P,Sc	S,L,Pr,R,P,Sc	M,Pr,R,Sh,C,P	S,L,Pr,R,Sh,C,P	M,Pr,R,Sh,C,Sc	S,L,Pr,R,Sh,C,Sc
L,P	M,L,P	L,Sc	M,L,Sh	L,Pr,R,P,Sc	M,L,Pr,R,P,Sc	L,Pr,R,Sh,C,P	M,L,Pr,R,Sh,C,P	L,Pr,R,Sh,C,Sc	M,L,Pr,R,Sh,C,Sc
S,Pr,R	S,M,Pr,R	S,Pr,Sh	S,M,Pr,Sh	S,Pr,C	S,M,Pr,C	S,Pr,P	S,M,Pr,P	S,Pr,Sc	S,M,Pr,Sc
M,Pr,R	S,L,Pr,R	M,Pr,Sh	S,L,Pr,Sh	M,Pr,C	S,L,Pr,C	M,Pr,P	S,L,Pr,P	M,Pr,Sc	S,L,Pr,Sc
L,Pr,R	M,L,Pr,R	L,Pr,Sh	M,L,Pr,Sh	L,Pr,C	M,L,Pr,C	L,Pr,P	M,L,Pr,P	L,Pr,Sc	M,L,Pr,Sc
S,Pu,R	S,M,Pu,R	S,Pu,Sh	S,M,Pu,Sh	S,Pu,C	S,M,Pu,C	S,Pu,P	S,M,Pu,P	S,Pu,Sc	S,M,Pu,Sc
M,Pu,R	S,L,Pu,R	M,Pu,Sh	S,L,Pu,Sh	M,Pu,C	S,L,Pu,C	M,Pu,P	S,L,Pu,P	M,Pu,Sc	S,L,Pu,Sc
L,Pu,R	M,L,Pu,R	L,Pu,Sh	M,L,Pu,Sh	L,Pu,C	M,L,Pu,C	L,Pu,P	M,L,Pu,P	L,Pu,Sc	M,L,Pu,Sc
S,Pr,R,Sh	S,M,Pr,R,Sh	S,Pr,R,C	S,M,Pr,R,C	S,Pr,R,P	S,M,Pr,R,P	S,Pr,R,Sc	S,M,Pr,R,Sc	S,Pr,R,Sh,C	S,M,Pr,R,Sh,C
M,Pr,R,Sh	S,L,Pr,R,Sh	M,Pr,R,C	S,L,Pr,R,C	M,Pr,R,P	S,L,Pr,R,P	M,Pr,R,Sc	S,L,Pr,R,Sc	M,Pr,R,Sh,C	S,L,Pr,R,Sh,C
L,Pr,R,Sh	M,L,Pr,R,Sh	L,Pr,R,C	M,L,Pr,R,C	L,Pr,R,P	M,L,Pr,R,P	L,Pr,R,Sc	M,L,Pr,R,Sc	L,Pr,R,Sh,C	M,L,Pr,R,Sh,C
S,Pr,R,Sh,P	S,M,Pr,R,Sh,P	S,Pr,R,Sh,Sc	S,M,Pr,R,Sh,Sc	S,Pu,R,P	S,M,Pu,R,P	S,Pu,R,Sc	S,M,Pu,R,Sc	S,Pu,R,Sh,C	S,M,Pu,R,Sh,C
M,Pr,R,Sh,P	S,L,Pr,R,Sh,P	M,Pr,R,Sh,Sc	S,L,Pr,R,Sh,Sc	M,Pu,R,P	S,L,Pu,R,P	M,Pu,R,Sc	S,L,Pu,R,Sc	M,Pu,R,Sh,C	S,L,Pu,R,Sh,C
L,Pr,R,Sh,P	M,L,Pr,R,Sh,P	L,Pr,R,Sh,Sc	M,L,Pr,R,Sh,Sc	L,Pu,R,P	M,L,Pu,R,P	L,Pu,R,Sc	M,L,Pu,R,Sc	L,Pu,R,Sh,C	M,L,Pu,R,Sh,C
S,Pr,R,C,P	S,M,Pr,R,C,P	S,Pr,R,C,Sc	S,M,Pr,R,C,Sc	S,Pu,R,P,Sc	S,M,Pu,R,P,Sc	S,Pu,R,Sh,C,P	S,M,Pu,R,Sh,C,P	S,Pu,R,Sh,C,Sc	S,M,Pu,R,Sh,C,Sc
M,Pr,R,C,P	S,L,Pr,R,C,P	M,Pr,R,C,Sc	S,L,Pr,R,C,Sc	M,Pu,R,P,Sc	S,L,Pu,R,P,Sc	M,Pu,R,Sh,C,P	S,L,Pu,R,Sh,C,P	M,Pu,R,Sh,C,Sc	S,L,Pu,R,Sh,C,Sc
L,Pr,R,C,P	M,L,Pr,R,C,P	L,Pr,R,C,Sc	M,L,Pr,R,C,Sc	L,Pu,R,P,Sc	M,L,Pu,R,P,Sc	L,Pu,R,Sh,C,P	M,L,Pu,R,Sh,C,P	L,Pu,R,Sh,C,Sc	M,L,Pu,R,Sh,C,Sc
S,Pr,R,Sh,P,Sc	S,M,Pr,R,Sh,P,Sc	S,Pr,R,C,P,Sc	S,M,Pr,R,C,P,Sc	S,Pr,Sh,Sc	S,M,Pr,Sh,Sc	S,Pr,Sh,C,P	S,M,Pr,Sh,C,P	S,Pr,Sh,C,Sc	S,M,Pr,Sh,C,Sc
M,Pr,R,Sh,P,Sc	S,L,Pr,R,Sh,P,Sc	M,Pr,R,C,P,Sc	S,L,Pr,R,C,P,Sc	M,Pr,Sh,Sc	S,L,Pr,Sh,Sc	M,Pr,Sh,C,P	S,L,Pr,Sh,C,P	M,Pr,Sh,C,Sc	S,L,Pr,Sh,C,Sc
L,Pr,R,Sh,P,Sc	M,L,Pr,R,Sh,P,Sc	L,Pr,R,C,P,Sc	M,L,Pr,R,C,P,Sc	L,Pr,Sh,Sc	M,L,Pr,Sh,Sc	L,Pr,Sh,C,P	M,L,Pr,Sh,C,P	L,Pr,Sh,C,Sc	M,L,Pr,Sh,C,Sc
S,Pu,R,Sh	S,M,Pu,R,Sh	S,Pu,R,C	S,M,Pu,R,C	S,Pu,Sh,Sc	S,M,Pu,Sh,Sc	S,Pu,Sh,C,P	S,M,Pu,Sh,C,P	S,Pu,Sh,C,Sc	S,M,Pu,Sh,C,Sc
M,Pu,R,Sh	S,L,Pu,R,Sh	M,Pu,R,C	S,L,Pu,R,C	M,Pu,Sh,Sc	S,L,Pu,Sh,Sc	M,Pu,Sh,C,P	S,L,Pu,Sh,C,P	M,Pu,Sh,C,Sc	S,L,Pu,Sh,C,Sc
L,Pu,R,Sh	M,L,Pu,R,Sh	L,Pu,R,C	M,L,Pu,R,C	L,Pu,Sh,Sc	M,L,Pu,Sh,Sc	L,Pu,Sh,C,P	M,L,Pu,Sh,C,P	L,Pu,Sh,C,Sc	M,L,Pu,Sh,C,Sc
S,Pu,R,Sh,P	S,M,Pu,R,Sh,P	S,Pu,R,Sh,Sc	S,M,Pu,R,Sh,Sc	S,Pu,C,P,Sc	S,M,Pu,C,P,Sc	S,Pr,P,Sc	S,M,Pr,P,Sc	S,Pu,P,Sc	S,M,Pu,P,Sc
M,Pu,R,Sh,P	S,L,Pu,R,Sh,P	M,Pu,R,Sh,Sc	S,L,Pu,R,Sh,Sc	M,Pu,C,P,Sc	S,L,Pu,C,P,Sc	M,Pr,P,Sc	S,L,Pr,P,Sc	M,Pu,P,Sc	S,L,Pu,P,Sc
L,Pu,R,Sh,P	M,L,Pu,R,Sh,P	L,Pu,R,Sh,Sc	M,L,Pu,R,Sh,Sc	L,Pu,C,P,Sc	M,L,Pu,C,P,Sc	L,Pr,P,Sc	M,L,Pr,P,Sc	L,Pu,P,Sc	M,L,Pu,P,Sc
S,Pu,R,C,P	S,M,Pu,R,C,P	S,Pu,R,C,Sc	S,M,Pu,R,C,Sc	S,Pr,C,P,Sc	S,M,Pr,C,P,Sc	S,Pu,C,P	S,M,Pu,C,P	S,Pu,C,Sc	S,M,Pu,C,Sc
M,Pu,R,C,P	S,L,Pu,R,C,P	M,Pu,R,C,Sc	S,L,Pu,R,C,Sc	M,Pr,C,P,Sc	S,L,Pr,C,P,Sc	M,Pu,C,P	S,L,Pu,C,P	M,Pu,C,Sc	S,L,Pu,C,Sc
L,Pu,R,C,P	M,L,Pu,R,C,P	L,Pu,R,C,Sc	M,L,Pu,R,C,Sc	L,Pr,C,P,Sc	M,L,Pr,C,P,Sc	L,Pu,C,P	M,L,Pu,C,P	L,Pu,C,Sc	M,L,Pu,C,Sc
S,Pu,R,Sh,P,Sc	S,M,Pu,R,Sh,P,Sc	S,Pu,R,C,P,Sc	S,M,Pu,R,C,P,Sc	Pr,Sc	Pr,R,Sh,C,P	Pr,R,Sh,C,Sc	Pr,R,Sh,C,Sc	Pr,R,Sh,P,Sc	Pr,R,C,P,Sc
M,Pu,R,Sh,P,Sc	S,L,Pu,R,Sh,P,Sc	M,Pu,R,C,P,Sc	S,L,Pu,R,C,P,Sc	Pu,Sc	Pu,R,Sh,C,P	Pu,R,Sh,C,Sc	Pu,R,Sh,C,Sc	Pu,R,Sh,P,Sc	Pu,R,C,P,Sc
L,Pu,R,Sh,P,Sc	M,L,Pu,R,Sh,P,Sc	L,Pu,R,C,P,Sc	M,L,Pu,R,C,P,Sc	Pr,R,Sh,C	Pr,R,Sh,P	Pr,R,Sh,Sc	Pr,R,C,P	Pr,R,C,Sc	Pr,R,P,Sc
S,Pr,Sh,C	S,M,Pr,Sh,C	S,Pr,Sh,P	S,M,Pr,Sh,P	Pu,R,Sh,C	Pu,R,Sh,P	Pu,R,Sh,Sc	Pu,R,C,P	Pu,R,C,Sc	Pu,R,P,Sc
M,Pr,Sh,C	S,L,Pr,Sh,C	M,Pr,Sh,P	S,L,Pr,Sh,P	Pr,Sh,C,Sc	Pr,Sh,P,Sc	Pr,Sh,C,P,Sc			
L,Pr,Sh,C	M,L,Pr,Sh,C	L,Pr,Sh,P	M,L,Pr,Sh,P	Pu,Sh,C,Sc	Pu,Sh,P,Sc	Pu,Sh,C,P,Sc			
S,Pr,Sh,P,Sc	S,M,Pr,Sh,P,Sc	S,Pr,Sh,C,P,Sc	S,M,Pr,Sh,C,P,Sc	Pr,C,P	Pr,C,Sc	Pr,C,P,Sc	Pu,C,P	Pu,C,Sc	Pu,C,P,Sc
M,Pr,Sh,P,Sc	S,L,Pr,Sh,P,Sc	M,Pr,Sh,C,P,Sc	S,L,Pr,Sh,C,P,Sc	Pr,Sh,C	Pr,Sh,P	Pr,Sh,Sc	Pr,Sh,C,P		
L,Pr,Sh,P,Sc	M,L,Pr,Sh,P,Sc	L,Pr,Sh,C,P,Sc	M,L,Pr,Sh,C,P,Sc	Pu,Sh,C	Pu,Sh,P	Pu,Sh,Sc	Pu,Sh,C,P		
S,Pu,Sh,C	S,M,Pu,Sh,C	S,Pu,Sh,P	S,M,Pu,Sh,P	Pr,R	Pr,Sh	Pr,C	Pr,P		
M,Pu,Sh,C	S,L,Pu,Sh,C	M,Pu,Sh,P	S,L,Pu,Sh,P	Pu,R	Pu,Sh	Pu,C	Pu,P		
L,Pu,Sh,C	M,L,Pu,Sh,C	L,Pu,Sh,P	M,L,Pu,Sh,P	Pr,R,Sh	Pr,R,C	Pr,R,P	Pr,R,Sc		
S,Pu,Sh,P,Sc	S,M,Pu,Sh,P,Sc	S,Pu,Sh,C,P,Sc	S,M,Pu,Sh,C,P,Sc	Pu,R,Sh	[illegible]		[illegible]		
M,Pu,Sh,P,Sc	S,L,Pu,Sh,P,Sc	M,Pu,Sh,C,P,Sc	S,L,Pu,Sh,C,P,Sc						
L,Pu,Sh,P,Sc	M,L,Pu,Sh,P,Sc	L,Pu,Sh,C,P,Sc	M,L,Pu,Sh,C,P,Sc						
S,Pr,C,P	S,M,Pr,C,P	S,Pr,C,Sc	S,M,Pr,C,Sc						
M,Pr,C,P	S,L,Pr,C,P	M,Pr,C,Sc	S,L,Pr,C,Sc						
L,Pr,C,P	M,L,Pr,C,P	L,Pr,C,Sc	M,L,Pr,C,Sc						

This cost data sub-group contains data for projects with the characteristic of : (1) school, carpark, commercial center and social community building, (2) small and large CFA and (3) Public client

Data were collected for 400 Hong Kong Private Sector and Public Sector (Housing Authority and Architectural Services Department) projects for analysing. Information regarding (1) floor area (CFA), (2) type of building, (3) type of client and (4) cost per HK$/m2 (1000) of the collected project data are recorded. Table 6 (attached in appendix) illustrates how the information is recorded and the ways of sub-grouping the cost data. In Table 6, the projects characteristics are recorded in Col. 2 and the project cost is recorded in Col.3. According to its characteristics, the project cost is pooled into its own group. For example, Project P1, a residential building (R) with private client (Pr) and small CFA (S), the cost is pooled under 4 categories: (1) CFA sub-groups in Col.4 (S), 7 (SM) and 8 (SL), (2) client type sub-groups in Col.10 (Pr), (3) building type sub-groups in C12 (R) and (4) a combination of sub-groups in Col. C17 (S, R, Pr), 22 (S, M, S, Pu), 23 (S, L, Pu, Sh, P) and C25 (All). Col. 4 therefore comprises the base-group of data for all small CFA buildings, while Col.17 is the base-group of data for small CFA residential buildings with a private client. The object is to find which pooling arrangement is the best for this project. As each project data collected contains not only one sub-group characteristic (e.g. CFA, Client Type and Project Type), but possesses combined characteristic (e.g. for project 1, it is a project with small CFA, residential building with private client). Therefore, the dataset of combining of the sub-groups can provide the most relevant data for forecasting.

PROPOSED METHOD TO IDENTIFY THE BEST DATA POOL TO BE USED FOR EARLY STAGE COST FORECASTS

The proposed method comprises three stages: (1) calculation of mean value forecast for the target-project via different base-groups, (2) measuring the error of ensuing forecasts and (3) identifying the best base-group by the value of the mean square error. The details are:

MEAN VALUE FORECAST FOR TARGET-PROJECT VIA DIFFERENT BASE-GROUPS

Base-groups can be divided into three types: (1) those with exactly the same characteristics as the target (e.g., for a target-project with a small CFA, the base-group comprises only those projects with a small CFA), (2) those with totally different characteristics to the target (e.g., for a target-project with a small CFA, the base-group comprises only those projects that do *not* have a small CFA), (3) those with both the same and different characteristics to the target (e.g., for a target-project with a small CFA, the base-group comprises those projects with a small CFA and medium GFA). Let X and Y be independent random variables representing the data with the same and different characteristics, respectively, to the target. X contains observations $x_1, x_2, \ldots, x_n$ and Y contains observations $y_1, y_2, \ldots, y_m$. The forecast is needed of the future, as yet unknown observation x_{n+1}. That is, letting $\hat{x}_{n+1} =$ forecast cost of the target-project,

Type I: using the base-group with exactly the same characteristic as the target

$$\hat{x}_{n+1} = \overline{x} \quad (1)$$

Where $\overline{x}$ = mean of the base-group with same characteristic

Type II: using the base-group with different characteristic to the target

$$\hat{x}_{n+1} = \overline{Y} \quad (2)$$

Where $\overline{Y}$ = mean of the base-group with different characteristic

Type III: using the base-group with both the same and different characteristics to the target

$$\hat{x}_{n+1} = \frac{n\overline{x} + m\overline{y}}{m+n} \quad (3)$$

Assume $\dfrac{n\overline{x} + m\overline{y}}{m+n} = \overline{Z}$ (4)

Where $\overline{z}$ = mean of the source data group with less similar characteristic

m = Nos. of data for *Y* data source group

n = Nos. of data for *X* data source group

FORECAST ERROR BETWEEN THE TARGET PROJECT AND THE MEAN FORECAST

The mean square error (MSQ) is used to measure the accuracy of the forecast construction cost. To simulate the forecast error, the out of sample MSQ is used. Therefore:

For **Type I**, the simulated forecast error is:

$$MSQ_{(I)} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}_i\right)^2 \quad (5)$$

Where $\overline{x}_i$ denotes the mean of the $x_1, x_2, x_3, \ldots\ldots, x_n$ observations excluding the *i*th observation

For **Type II**, the simulated forecast error is :

$$MSQ_{(II)} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{y}\right)^2 \quad (6)$$

Where $\overline{y}$ denotes the mean of the $y_1, y_2, y_3, \ldots\ldots, y_n$ observations.

For **Type III**, the simulated forecast error is:

$$MSQ_{(III)} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{z}_i\right)^2 \quad (7)$$

Where $\overline{z}_i$ denotes the mean of the $z_1, z_2, z_3, \ldots\ldots, z_n$ observations excluding the *i*th observation

IDENTIFYING THE BEST DATA-POOLING ARRANGEMENT

To identify the best data-pooling arrangement, simulated forecasts are made for each and every project in the target-group for a base-group composed by candidate data-pooling arrangement. These are then subjected to an error test and the mean square error values are recorded. The base-group that generates the least mean square error value is the best base-group for the target project and the associated data-pooling arrangement is therefore the best data-pooling arrangement.

TRIAL ANALYSIS ON THE PROPOSED METHOD

For demonstration purposes, a trial analysis of the CFA group is conducted. 145 projects are pooled into sub-groups according to construction floor area (CFA) as described in Table 2. The total numbers of permutations are (1) Small CFA Group, (2) Medium CFA Group, (3) Large CFA Group, (4) Small and Medium CFA Group, (5) Small and Large CFA Group, and (6) Medium and Large CFA Group. The mean square error values for each pooling are tabulated in Table 7.

Table 7: **Mean square error result for CFA sub-group**

The least MSQ in CFA Sub-group

Project-group	**CFA base-group**					
	Small	**Medium**	**Large**	**Small & Medium**	**Small & Large**	**Medium & Large**
Small	4.664	4.791	6.297	1.239	1.339	1.386
Medium	3.684	3.590	4.457	0.795	0. 868	0.764
Large	7.730	6.462	6.081	1.634	1.589	1.739

From the above analysis result, it is noted that the traditional way of forecast using exactly the same characteristic data source (Type I) cannot generate the most accurate result. For example, small CFA target projects have a simulated mean square error of 4.664 when using the only small CFA base-group, while using the combined small and medium CFA base-group, the lowest mean square value of 1.239 is obtained. This indicates that the small and medium CFA base-group provides the best pooling arrangement for forecasting small CFA projects. Similarly, pooling the medium and large CFA projects provides the forecast for medium CFA projects (MSQ=0.764), while, somewhat surprisingly, pooling the small and large CFA projects provides the best forecast for large CFA projects.

CONCLUSION

In extreme situations where only such basic information as project type and size is known concerning a new (target) project, forecasters have to resort to either using the price of a very similar project or mean price of a (base) group of projects. As is known, enlarging the size of the base-group lessens the variability of the forecast made this way but, at the same time, also lessens its appropriateness (homogeneity). An empirical method is proposed for identifying the base-group that provides the best trade-off of these opposing features. A trial analysis is described involving the analysis of 145 actual Hong Kong construction projects in terms of small, medium and large floor area (CFA). This demonstrates that using the historical data with the same characteristics may not generate the best forecasts in the early stage, with the small and medium CFA projects providing the best forecasts for small CFA targets, medium and large CFA projects providing the best forecasts for medium CFA targets, and small and large CFA projects providing the best forecasts for large CFA targets.

The trial analysis provided above is not comprehensive, however. The analysis only concentrated on one grouping category (CFA grouping) with 145 projects. The method should easily extend to other possible grouping categories and further sub-division of the data groups as mentioned in Table 2 to Table 5 (i.e. CFA group, Client Type, Building Type and the permutation of different sub-groups). To improve the result of confident, data from more projects are proposed to collect for further analysis.

REFERENCES

Morrison, N.A.D. (1984) The accuracy of quantity surveyors’ cost estimating. Construction Management and Economics,2 57-75.

Beeston, D.T. (1975) One statistician’s view of estimateing Chartered Surveyor: Building and Quantity Surveying Quarterly, 49-54.

Flanagan, (1980) Tender price and time prediction for construction work, PhD Thesis, University of Aston in Birmingham.

Flanagan, R. and Norman, G (1983) The accuracy and monitoring of quantity surveyors’ price forecasting for building work. Construction Management and Economics, 157-180.

Skitmore, M. (2001) Raftery curves for tender price forecasting:empirical probabilities and pooling. Financial Management of Property and Construction, 6

Skitmore, R.M. (1991) Early stage construction price forecasting; a review of performance, The Royal Institution of Chartered Surveyors.

Skitmore, R.M. (2005) Data pooling for mean value forecasting, unpublished paper, Queensland University of Technology.

Skitmore, R.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, Department of Civil Engineering, The University of Salford.

Skitmore, R.M. and Tan, S.H. (1988) Factors affecting the accuracy of engineers’ estimates, Transactions 10th International Cost Engineering Congress, The American Association of Cost Engineers, paper B-3.

Tan, S. H. (1988) An investigation into the accuracy of cost estimates during the design stages of construction projects, Department of Civil Engineering, The University of Salford