CONTRIBUTION TO THE AMELIORATION OF THE ESTIMATION METHOD OF CONSTRUCTION COSTS’ MASTERING IN DEVELOPING COUNTRIES

P. Louzolo-Kimbembé* and C. Pettang**
* Ecole Normale Supérieure, Département des Sciences Exactes, B.P. 237 – Brazzaville / Congo
** Ecole Nationale Supérieure Polytechnique, Département de Génie Civil, B.P. 8390 – Younde / Cameroun

(Received 15 November 2003; Accepted 30 June 2006)

ABSTRACT

In most developing countries the construction costs seem too high to the majority of the population made up of families with very low incomes. One of the means susceptible to contribute to the reduction of costs is its proper mastering. Cost estimation based on matrix approach leads to the description of the principal components of construction cost in a detailed manner. However this model does not take into consideration cost fluctuations through the uncertainty. This article presents an amelioration of the matrix approach using multiple linear regression. We try to show that the estimation of the total construction cost can be inside a tolerance margin known in advance. Also with the improvement of the estimation method, we want to arrive to the mastering of the total construction cost thanks to a forecast allowing being located in the cost reduction zone.

1. INTRODUCTION

Faced with the rapid and accelerated urbanisation of Developing Countries (DC), mostly due to demographic explosion or rural exodus [1,2], lodging problem often crops up as the greatest challenge to most governments. Unfortunately the population affected with this problem is the population with a low standard of living as they constitute the majority in urban area [3]. The failure of construction policies of social lodgings by the general public, due to diverse reasons with financial difficulties being acute [4], this has gone a long way to delay the efforts put by the informal sector in the domain of construction. Today households with modest incomes have turned to self-production (or self-construction, or self-promotion) as the last resort. The consequences being the presence of low quality lodgings seen in our urban centres; financial insufficiency being one of the major causes.

The greatest question remains how to resolve this antinomic situation of households with low incomes who should live in housings with acceptable conditions of which the cost of constructing a decent lodging is out of their reach? The most plausible response to this question will be to react so as to reduce substantially the construction cost. Many solutions had been proposed as to this amongst which we have the promotion of construction out of local material [5]. This solution has not provided the expected results for lodging, problems still remain acute.

Here we will envisage construction cost mastering as one of the ameliorating cost estimation methods which occupies the principal part of this literature [6,7,8]. These different usual estimation methods such as unit method, cube method, superficial or floor area method, storey-enclosure method, etc. [8], however hides other important parameters such as construction’s data line, labour of the other trades involved, etc. Other classic estimation methods well developed, such as the cost index method and the construction complexity method [9], or the estimation method based on the work breakdown structure [7], are difficult to transpose in the majority of the DC. Indeed, much of indexes and of factors used in these methods must be available and especially reliable, which is still not the case in most of DC. The consequences of all these gaps are illustrated in the absence of precision of the estimated assessment of the total construction cost.

With the aim of achieving a good cost mastering, it is necessary to describe the principal components in a more detailed manner. Only the matrix approach advocates the decomposition of total construction cost into three fundamental elements: material, labour, management means [10]. This leads us to the following expression:

\[ C_T = C_{Ma} + C_{La} + C_{Mg} \]  \hspace{1cm} (1)

where \( C_T \) is the total construction cost; \( C_{Ma} \) is the material cost; \( C_{La} \) is the labour cost; and \( C_{Mg} \) is the management cost.

Moreover a structure can undergo a number of sub-structural decomposition depending on the realization compatibility rules of the constitutive
elements of each sub-structure. Thus expression (1) could be put in the form:

\[ C_T = \sum_{k=1}^{s} \left( C_{Ma,SOk} + C_{La,SOk} + C_{Mg,SOk} \right) \quad (2) \]

where \( C_{Ma,SOk} \) is the material cost necessary for the realization of the sub-structure \( SOk \); \( C_{La,SOk} \) is the labour cost of all works interfering with the realization of sub-structure \( SOk \); \( C_{Mg,SOk} \) is the management cost necessary for the realization expenditure of the sub-structure \( SOk \); and \( s \) is the total number of sub-structures.

The expression (2) can still be written by introducing the costs by materials, labour and management expenditure. Thus:

\[ C_T = \sum_{k=1}^{s} \left[ \sum_{j=1}^{p} m_k^j + \sum_{h=1}^{q} w_k^h + \sum_{l=1}^{r} g_k^l \right] \quad (3) \]

where \( p \) is the number of elementary components (materials); \( q \) is the number of trade intervening in all the structure; \( r \) is the quantity of management means necessary for the realization of all the structure; \( m_k^j \) is the material cost of material mat \( j \) necessary for the realization of the sub-structure \( SOk \); \( w_k^h \) is the labour cost of all the workers of the trade trd \( h \) for the realization of the sub-structure \( SOk \); and \( g_k^l \) is the cost of the expenditure exp \( l \) necessary for the realization of the sub-structure \( SOk \).

The matrix approach presents a great advantage to self-producers (or self-constructors, or self-promoters). Indeed this approach permits us to have a rapid knowledge of the principal components that contribute in the constitution of construction cost, thus facilitating the detailed estimation of the project. Nevertheless the matrix approach does not take into account instability due to the realization time limit of different sub-structures and also with the incidence of economic perturbations. This weakness is related to the determinist character of the matrix approach. It is therefore necessary to foresee eventual fluctuations in the construction cost estimation. This imperceptible share could be introduced inside a tolerance margin. Indeed the incorporation of uncertainty in the planning and the management of a project is very important [11]. We would have to estimate and to control uncertainty, because the uncertainty contained in the estimation is information as important as the estimated value itself [12]. Our proposition consists of applying the multiple linear regression model in order to introduce corrective parameters in calculating costs.

This article is made up of four sections.

In section 2, we present our construction cost estimation model using the multiple linear regression as well as the method adopted. In section 3, we present a case study for the implementation of the model. Lastly in section 4 we advance our conclusion on the pertinence of the model in the context of construction costs’ mastering in DC by introducing the concept of “cost reduction zone”.

2. MODEL AND METHOD

2.1 Model

From the matrix approach we can write the relation:

\[ C_T = f(C_{Ma}, C_{La}, C_{Mg}) \quad (4) \]

where \( C_T \) represents the total construction cost.

The elements \( C_{Ma}, C_{La} \) and \( C_{Mg} \) enclose the following variables:

- material cost \( C_{Ma}: c_1, c_2, \ldots, c_j, \ldots, c_p; j = 1, \ldots, p \)
- labour cost \( C_{La}: c_{p+1}, c_{p+2}, \ldots, c_j, \ldots, c_q; j = p+1, \ldots, q \)
- management cost \( C_{Mg}: c_{q+1}, c_{q+2}, \ldots, c_j, \ldots, c_r; j = q+1, \ldots, r \)

It consists of considering a dependent variable (to be explained) \( C_T \) and a certain number of regressors or independent variables (explanatory) \( c_1, c_2, \ldots, c_j, \ldots, c_r \). Another way of writing the relation between \( C_T \) and the \( c_j \) is by using the multiple regression model of type [13,14]:

\[ C_T = \alpha_0 + \alpha_1 c_1 + \alpha_2 c_2 + \ldots + \alpha_r c_r + \varepsilon \quad (5) \]

where the variable \( \varepsilon \) represents the individual behaviour (random error).

Our aim is to express the total construction cost as a linear combination of independent variables \( r \). Let’s consider data from a statistic sample \( n \) (\( n \): number of constructed lodgings). The multiple linear regression model consist of supposing [15,16] that for all lodgings \( i \) (\( i = 1, \ldots, n \)):

\[ C_{Ti} = \alpha_0 + \sum_{j=1}^{r} \alpha_j c_{ij} + \varepsilon_i \quad (6) \]

where \( C_{Ti} \) is the total construction cost of lodging \( i \); \( \alpha_0 \) is the constant term; \( \alpha_j \) is the parameters, or regression coefficients; \( c_{ij} \) is \( c_j \) values (\( j = 1, \ldots, r \)) for lodging \( i \); and \( \varepsilon_i \) is the value of the random variable \( \varepsilon \) for lodging \( i \).

The introduction of matrix notation leads us to replace the \( n \) equations (6) by only a single matrix equation [13,14,17]:

\[ C_T = C\alpha + \varepsilon \quad (7) \]
where $C_T$ is the (nx1) observation vector relative to the explained variable; $C$ is the $[nx(r+1)]$ matrix relative to explicative variables with an additional column of 1 which correspond to the constant parameter $\alpha_0$; $\alpha$ is the $[(r+1)x1]$ unknown parametric vector to be estimated; and $\varepsilon$ is the (nx1) error vector.

Important to note that $C_T$ and $C$ are raw data.

The problem consists of estimating the vector $\alpha$ through the estimator vector $A$. The vector of estimated values $K_T$ is therefore:

$$K_T = C A \quad (8)$$

Therefore equation (8) represents the matrix of values predicted by the model. $K_T$ is a random vector as $A$ is random vector. The model’s vector of residues (or estimated deviation) is obtained by writing:

$$e = C_T - K_T \quad (9)$$

The ultimate step of the regression is the control of the quality of equation (8). This step turns out important as it permits us to know the variation of the explained variable $C_T$ that can be attributed to the regression [17]. However, the coefficients (parameters) of regression create a certain number of questions such as their non-nullity or their negligibility which would have had as a consequence the elimination of explicative variables to which they are affected. The non-nullity permits us to know if effectively there exist an association link between the explained variable and the explicative variables. The study of the relative weight of each variable of the model is of double interest: either to ameliorate, either to simplify.

Thanks to hypothesis tests on the model’s parameters of linear regression and of the estimations through the confidence interval, we would be able to reduce the number of independent variables through the elimination of non significant parameters in the presence of the complete model. This permitting finally to express the model simpler with less explicative variables. Indeed the objective of a predictive model is to emphasize on the quality of the estimators who must minimize a quadratic error. This resulted in seeking models with a restricted number of independent variables. Thus a good model is not anymore that which explains best the data within the meaning of a minimal deviance in comparison with a significant number of variables which can introduce collinearities. A good model is that which leads to the most reliable predictions [14].

### 2.2 Method

Let’s consider raw data from a sample of $n$ constructions of the same type. It is important to note that the real cost of lodgings is that established at the end of construction. All the lodgings that shall be retained must have been constructed almost within the same period in order to limit the eventual fluctuations of material price or the management expenditure. Independent variables are of three types: manpower or labour, materials, management means. Concerning materials the structure decomposition diagram below (Fig. 1) provides us the possibility to come out with the elementary constituents necessary in a construction. We can thus proceed in regrouping the same materials contained in the elementary parts of sub-structures. This leads to sensible diminution of the number of starting variables.

The principal sub-structures are:

- $SO_1$: Escavation and setting;
- $SO_2$: Foundation and ground slab;
- $SO_3$: Elevation;
- $SO_4$: Roofing;
- $SO_5$: Ceiling;
- $SO_6$: Carpentry;
- $SO_7$: Sanitary plumbing;
- $SO_8$: Electricity;
- $SO_9$: Plastering;
- $SO_{10}$: Face work and tiling;
- $SO_{11}$: Painting.

For instance, the elementary parts contained in the sub-structure $SO_2$ are: lean concreting, footings, basement wall, ground chaining and column footings. We can now open up the elementary constituents (basic materials) common to these elementary parts: cement and sand; then we regroup. Thus, in proceeding we still have the elementary constituents.

A software such as SAS (Statistical Analysis System) provides the means to carry out automatic sorting of regressors, bearing without any problems up to 15 regressors. This covers most parts of applications [18]. This limit imposed by the software capacity obliges the total number of independent variables not to exceed 15. Let us remark that in our work we use, for the statistical data processing, the software SPSS (Statistical Package for Social Science) which is as powerful as SAS, this for reasons of availability.

Thus to reduce the number of variables we could neglect the elementary constituents of less importance amongst the materials. The trades shall be regrouped in a multiskilled labour sense (e.g. carpentry-framework, builder-tiller, etc.). Indeed the multiskilled labour could be a strategy of labour which would help to reduce the indirect work costs.
to improve the productivity and to reduce the renewal of the personnel [19,20]. Depending on the means of management, we shall retain only significant expenses. The calculation of independent variables $c_{ij}$ shall be done from the matrix model.

**2.2.1 Determination of the material cost matrix**

Let $M$ be the matrix representing the material cost. It is defined as the product of the material quantity matrix $Q$ with the unit price matrix $U$ associated with different materials [10]. The matrix $Q$ is arrayed in such a way that the rows represent sub-structures and the columns represent materials. The matrix $M$ is written as:

$$M = QU = \{ q_{jk} U_k \} = \{ m_{jk} \}$$  \hspace{1cm} (10)

where $q_{jk}$ and $m_{jk}$ represent respectively the quantity and the material cost of material mat$_j$ necessary for the realization of the sub-structure $SO_k$, with $k = 1, \ldots, s$ and $j = 1, \ldots, p$; while $U_k$ is the unit price for material mat$_k$. Note that in order to eliminate redundant terms, it is necessary to consider the hypothesis of the independence of unit prices of $p$ elementary constituents. In this case $U$ becomes a diagonal matrix. We will write: $U_k = u_j \delta_{kj}$, where $\delta_{kj}$ is the Kronecker delta symbol. The matrix $M$ can now be expressed as:
N.B. If the material mat\textsubscript{i} is not used in the sub-structure SO\textsubscript{k}, then \( q_{i,k} = 0 \).

### 2.2.2 Determination of the labour cost matrix

The quantifiable variables that enter the calculation of labour cost are: trades, total number in each trade, daily revenue and the realization time of sub-structures [21]. If we denote \( N, R, D \) the matrices representing respectively the total number per trade, daily revenues and the realization time of sub-structures, then the labour cost matrix \( W \) could be written as the product of the three matrices:

\[
W = (NR)D \quad (11)
\]

In order to eliminate the redundant terms in the matrix \( W \), we pose the following hypothesis:

*The same trade receive uniform revenue whatever the sub-structure in which it intervenes.*

Consequence: \( R \) becomes a diagonal matrix. It is a square matrix whose order equals the number of trades engaged in the construction. The product \( (NR) \) represents the daily revenues matrix of all the intervening trades given by each sub-structure. Finally the general term of the matrix \( W \) can be written as:

\[
\{ w_{j,k} \} = \{ n_{j,k} R_j^k d_{j,k} \} \quad (12)
\]

where \( w_{j,k} \) is the labour cost of all the workers of trade \( \text{tr}_d \), for the realization of the sub-structure \( \text{SO}_k \); \( n_{j,k} \) is the total number of workers interfering in trade \( \text{tr}_d \) for the realization of sub-structure \( \text{SO}_k \); \( R_j^k \) is the daily revenue of a trade \( \text{tr}_d \), with \( R_j^k = r_j^k \delta_{jk} \); here \( j = p+1, p+2, ..., q \); \( d_{j,k} \) is the realization time of sub-structure \( \text{SO}_k \).

If a given trade does not intervene in the realization of sub-structure \( \text{SO}_k \), then \( n_{j,k} = 0 \).

### 2.2.3 Determination of the management cost matrix

We understand as construction management cost the total expenditure engaged in the execution of works in a construction site. Let \( G \) be the matrix representing the management cost. We write:

\[
G = VE \quad (13)
\]

where \( V \) and \( E \) are matrices denoting respectively the volume of charges (petrol, maintenance, etc.) and the elementary expenditure supposed independent. We then have:

\[
G = \{ v_{j,k} E_j^k \} = \{ g_{j,k} \} \quad (14)
\]

where \( v_{j,k} \) is the volume of proper charge char\textsubscript{j} of sub-structure \( \text{SO}_k \); \( E_j^k \) is the elementary expenditure relative to the charge char\textsubscript{j}; with \( E_j^k = e_j^k \delta_{jk} \); \( g_{j,k} \) is the amount of expenditure \( \text{exp}_j \) necessary for the execution of works linked to sub-structure \( \text{SO}_k \); and \( j = q+1, q+2, ..., r \).

### 2.2.4 Determination of the independent variables

#### 2.2.4.1 Variables relative to the material cost

For a given lodging \( i \), the cost \( c_{ij} \) of material \( \text{mat}_j \) necessary for all the structure is:

\[
c_{ij} = \sum_{k=1}^{s} (m^k_{ij}) \quad (15)
\]

with \( j = 1, ..., p \) and \( i = 1, ..., n \).

Let for the lodging \( i \):
Materials → mat₁ mat₂ ... matₚ
SO₁ → m¹₁ m¹₂ ... m¹ₙ
SO₂ → m²₁ m²₂ ... m²ₙ
... ...
SOₖ → mₖ¹ mₖ₂ ... mₖₙ
SOₗ → mₗ¹ mₗ₂ ... mₗₙ
Sub-structures ↑ ↑ ↑
c₁₁ c₁₂ ... c₁ₚ
... ...
cₙ₁ cₙ₂ ... cₙₚ

Thus the total cost of material matᵢ relative to the lodging i is given by:

\[ cᵢ = (mᵢ₁)i + (mᵢ₂)i + ... + (mᵢₙ)i \] (16)

2.2.4.2 Variables relative to the labour cost

Let’s have the lodging i. The cost labour cᵢⱼ linked to the trade tradⱼ for the realization of all the structure is given by the expression:

\[ cᵢⱼ = \sum_{k=1}^{s} (wⱼk)i \] (17)

with j = p+1,...,q and i = 1, ..., n.

2.2.4.3 Variables relative to the management cost

For a given lodging i, the cost cᵢⱼ of management means proper to a type of expenditure expⱼ for the structure is expressed by the relation:

\[ cᵢⱼ = \sum_{k=1}^{s} (gⱼk)i \] (18)

2.2.5 Construction of data base

With a sample sufficiently large, size n observations (constructions), a dependent variable and r regressors \( c_j \) (j = 1,..., r), we wish to establish a relation between the total construction cost \( C_Ti \) and the r variables \( c_j \) where:

- \([c₁₁ ; c₁ₚ] :\) construction material cost for lodging i;
- \([cᵢ,p+1 ; cᵢ,q] :\) labour cost for lodging i;
- \([cᵢ,q+1 ; cᵢ,r] :\) management cost for lodging i.

It is necessary firstly for us to evaluate the variables \( c_j \) for each construction as shown in 2.2.4. After calculation the data are listed in a table as in Table 1.

The estimated total construction cost for lodging can be drawn from the equation (8). It is given by the following expression:

\[ K_Ti = A₀ + \sum_{j=1}^{p} A_j cᵢⱼ + \sum_{j=p+1}^{q} A_j cᵢⱼ + \sum_{j=q+1}^{r} A_j cᵢⱼ \]

\[ C_{Ma} \ C_{La} \ C_{Mg} \] (19)

According to relative weights of each variable \( c_j \), indicated by \( A_j \), certain independent variables could be neglected. That means that all the \( c_j \) do not appear in the expression of the estimated total construction cost. Only the dominating variables are considered. It is thus possible to estimate the total construction cost with an error margin known in advance, since the coefficients \( A₀ \) and \( A_j \) are defined with confidence intervals.

Table 1: Construction cost, material cost, labour cost and management cost

<table>
<thead>
<tr>
<th>Lodging i</th>
<th>Construction cost</th>
<th>Construction material cost</th>
<th>Labour cost</th>
<th>Management cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( C_{T1} )</td>
<td>( c₁₁ ) ( c₁₂ ) ... ( c₁ₚ )</td>
<td>( c₁,p+1 ) ( c₁,p+2 ) ... ( c₁,q )</td>
<td>( c₁,q+1 ) ( c₁,q+2 ) ... ( c₁,r )</td>
</tr>
<tr>
<td>1</td>
<td>( C_{T2} )</td>
<td>( c₁₁ ) ( c₁₂ ) ... ( c₁ₚ )</td>
<td>( c₁,p+1 ) ( c₁,p+2 ) ... ( c₁,q )</td>
<td>( c₁,q+1 ) ( c₁,q+2 ) ... ( c₁,r )</td>
</tr>
<tr>
<td>2</td>
<td>( ... )</td>
<td>( ... ) ( ... ) ... ( ... )</td>
<td>( ... ) ( ... ) ... ( ... )</td>
<td>( ... ) ( ... ) ... ( ... )</td>
</tr>
<tr>
<td>...</td>
<td>( ... )</td>
<td>( ... ) ( ... ) ... ( ... )</td>
<td>( ... ) ( ... ) ... ( ... )</td>
<td>( ... ) ( ... ) ... ( ... )</td>
</tr>
<tr>
<td>n</td>
<td>( C_{Tn} )</td>
<td>( c₁₁ ) ( c₁₂ ) ... ( c₁ₚ )</td>
<td>( c₁,p+1 ) ( c₁,p+2 ) ... ( c₁,q )</td>
<td>( c₁,q+1 ) ( c₁,q+2 ) ... ( c₁,r )</td>
</tr>
<tr>
<td></td>
<td>( ... )</td>
<td>( ... ) ( ... ) ... ( ... )</td>
<td>( ... ) ( ... ) ... ( ... )</td>
<td>( ... ) ( ... ) ... ( ... )</td>
</tr>
<tr>
<td></td>
<td>( ... )</td>
<td>( ... ) ( ... ) ... ( ... )</td>
<td>( ... ) ( ... ) ... ( ... )</td>
<td>( ... ) ( ... ) ... ( ... )</td>
</tr>
</tbody>
</table>

For example: \( c₁ \): cement cost; \( c₂ \): aggregate cost; \( ... \); \( cₚ \): maintenance cost, etc.
3. CASE STUDY: IMPLEMENTATION OF THE CONSTRUCTION COST ESTIMATION METHOD

To be valid the data on the basic components of construction, for the numerical determination of the coefficients $A_j$, must come from the analysis of relatively similar constructions. Thus for the implementation of the method, we considered only one family of constructions, namely the residential buildings in ground floor. Let us note however that the fact of considering only residences in ground floor restricts certainly the applications, but does not change anything with their fundamental direction. Moreover, the great majority of the self-producers choose constructions in ground floor because purchasing power to them very limited.

3.1 Presentation of the Construction Cost Data

The geographical field which enabled us to collect the data is the town of Brazzaville (Congo). We consulted 369 constructions files covering the period 2000-2003. Only 18 files (5% of the total), with costs higher than 10 millions F CFA (1Euro = 656 F CFA), presented complete and exploitable data. We treated the files so as to gather the components of the construction costs. We thus obtain 8 categories of building materials and 5 types of trade.

- Material cost:
  1. Mat-Masonry (matmas): cost of materials used in masonry work;
  2. Mat-Carpentry/Joinery (matcajo): cost of materials used in carpentry and joinery work;
  3. Mat-Plumbing (matplu): cost of materials used in plumbing work;
  4. Mat-Tiling/Facing (mattifa): cost of materials used in tiling and facing work;
  5. Mat-Roofing/Sealing (matrose): cost of materials used in roofing and sealing work;
  6. Mat-Painting (matpa): cost of materials used in painting work;
  7. Mat-Electricity (matel): cost of materials used in electricity work;
  8. Mat-Glaziery (matgla): cost of materials used in glaziery work.

- Labour cost:
  9. Lab-MTFS (labmtfs): labour cost representing the regrouping mason/tile-layer/frame worker/steel bender;
  10. Lab-CJG (labcjg): labour cost representing the regrouping carpenter/joiner/glazier;
  11. Lab-Plumber ((labplu): labour cost representing the trade plumber;
  12. Lab-Electrician (label): labour cost representing the trade electrician;
  13. Lab-Painter (labpa): labour cost representing the trade painter.

The table in Annexe 1 presents the summary of the results. We find the total construction cost $C_T$ by directly adding all the elements of a given line. We then calculated the contribution in weight of each element of the cost compared to the average total cost. This enabled us to determine the elements having a dominating weight among the material and labour cost.

For the prediction of the total construction cost ($C_T$), which represents the dependent variable, we have needs only for the independent variables most significant. We thus retained only those which contribute to 5% at least on average in the total cost. Finally the useful independent variables are: “matmas”, “matcajo”, “matplu”, “mattifa”, “matrose”, “matpa”, “matel”, “labmtfs”, “labcjg”.

3.2 Statistical Data Processing

For the determination of the linear regression statistics, we use the software SPSS, version 10.0.5 [22]. We chose the Stepwise procedure to analyze the data. Eight models were tested successively by SPSS. The analysis of the variance showed that the models obtained are significant, with sig. = 0.000. Let us note that the error of first type or significance (sig.) informs about the risk to be mistaken on the direction in the regression. If sig. < 0.05 we can conclude in favour of the existence of a linear regression model, with threshold 0.05 (with the significance threshold indicated by the statistics sig.). The regression coefficients are significant in all the models, apart from the constants with sig. > 0.05. We can thus consider that the nullity hypothesis of the coefficients can be rejected, except for the constants. We finally retained only the most significant model among the eight tested. In this model there remain only five independent variables. Table 2 presents the regression coefficients.

Several variables seem to present collinearity statistics which are not good enough, with a tolerance < 0.3 and a VIF (variance inflation factor) > 3.3. Indeed the limits prescribed for the tolerance and the VIF are: tolerance > 0.3 and VIF < 3.3 [23]. The collinearity diagnosis makes it possible to show if in a model the variables are collinear, i.e. if they are correlated between them. The absence of collinearity problems on the explanatory variables is an indication of good quality of the model. The choice of the only model selected was justified by the fact that this last presented less collinearity problems than the other models generated by the software. Table 3 gives an indication on the collinearity diagnoses.
Table 2: Regression coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Non-standardised Coefficients</th>
<th>t</th>
<th>Sig.</th>
<th>Confidence interval for B at 95%</th>
<th>Collinearity statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Standard deviation</td>
<td>Inferior limit</td>
<td>Superior limit</td>
<td>Tolerance</td>
</tr>
<tr>
<td>(Constant)</td>
<td>-491090,6</td>
<td>332290,21</td>
<td>-1,478</td>
<td>1215088,673</td>
<td>0,030</td>
</tr>
<tr>
<td>labmtfs</td>
<td>4,958</td>
<td>0,573</td>
<td>8,656</td>
<td>3,710</td>
<td>6,207</td>
</tr>
<tr>
<td>matrose</td>
<td>3,038</td>
<td>0,422</td>
<td>7,193</td>
<td>2,117</td>
<td>3,958</td>
</tr>
<tr>
<td>matmas</td>
<td>0,712</td>
<td>0,086</td>
<td>8,305</td>
<td>0,525</td>
<td>0,899</td>
</tr>
<tr>
<td>matcajo</td>
<td>0,594</td>
<td>0,118</td>
<td>5,017</td>
<td>0,336</td>
<td>0,852</td>
</tr>
<tr>
<td>matplu</td>
<td>2,109</td>
<td>0,636</td>
<td>3,314</td>
<td>0,722</td>
<td>3,495</td>
</tr>
</tbody>
</table>

The variable “labmtfs” is related to dimension 6 and very little with the others, “matrose” with dimension 5 and very little with the others, “matmas” with dimension 4 and very little with the others, “matcajo” with dimension 3 and very little with the others, “matplu” with dimension 6 and very little with the others. Thus each variable is correlated with a different dimension and only one, except for “labmtfs” and “matplu” which is related both to the same dimension 6, which is confirmed by their correlation coefficient (−0.685) which is highest in the model.

3.3 Expression of the Total Construction Cost

From the formula (19) and the numerical values of the regression coefficients given in the table above, we obtain:

\[
K_T = -5 \times 10^5 + 5 \times \text{labmtfs} + 3 \times \text{matrose} + 0.7 \times \text{matmas} + 0.6 \times \text{matcajo} + 2 \times \text{matplu}
\]

with \( A_0 = 5 \times 10^5; \ A_1 = 5; \ A_2 = 3; \ A_3 = 0.7; \ A_4 = 0.6; \ A_5 = 2. \)

The confidence intervals for the regression coefficients are respectively: \( A_0: [-1.2 \times 10^6; 2.3 \times 10^5]; \ A_1: [3.7; 6.2]; \ A_2: [2.1; 4.0]; \ A_3: [0.5; 0.9]; \ A_4: [0.3; 0.9]; \ A_5: [0.7; 3.5]. \) The construction cost can thus be calculated with a certain tolerance. Indeed we have the low estimation \( K_{Th} \) and the high estimation \( K_{Th} \) respectively given by the following expressions:

\[
K_{Th} = -1.2 \times 10^6 + 3.7 \times \text{labmtfs} + 2.1 \times \text{matrose} + 0.5 \times \text{matmas} + 0.3 \times \text{matcajo} + 0.7 \times \text{matplu}
\]

\[
K_{Th} = 2.3 \times 10^5 + 6.2 \times \text{labmtfs} + 4 \times \text{MATROSE} + 0.9 \times \text{MATMAS} + 0.9 \times \text{MATCAJO} + 3.5 \times \text{MATPLU}
\]

We thus have a cost margin such as:

\[
M = \frac{(K_{Th} - K_{Th})}{2}
\]

Finally the complete expression of the total construction cost will be written as:

\[
K_{Th} = -5 \times 10^5 + 5 \times \text{labmtfs} + 3 \times \text{matrose} + 0.7 \times \text{matmas} + 0.6 \times \text{matcajo} + 2 \times \text{matplu} \pm M
\]

3.4 Verification of the Construction Cost Estimation Model

a) Let us apply the formula (20) in the case of housing n°15 (see table in Annexe 1). We find the following estimated total cost:

\[
K_T = -5 \times 10^5 + 5 \times (2 940 000) + 3 \times (2 366 800) + 0.7 \times (4 479 780) + 0.6 \times (4 883 100) + 2 \times (2 242 980)
\]

\[
= 35 352 066 \text{ F CFA}
\]
The raw total cost being \( C_T = 35 \, 243 \, 706 \, \text{F CFA} \), we obtain a low value of the residue (formula (9)), \( e = 108 \, 360 \, \text{F CFA} \). On the other hand the corresponding margin \( M \) (formula (21)) remains rather large, we find \( M = 13 \, 140 \, 000 \, \text{F CFA} \).

\section*{b) Now let us consider a residential building in ground floor of useful area equal to 110 m\(^2\), built out of conventional materials [5]. The data on the costs of this construction [24] enabled us to calculate the various variables hereafter:

\begin{itemize}
  \item labmtfs = labor cost (mason; frame worker; tile-layer; steel bender)
    \[ = 408 \, 000 \, \text{F} + 21 \, 000 \, \text{F} + \]
    \[ 200000 \, \text{F} + 21 \, 000 \, \text{F} = 650 \, 000 \, \text{F CFA} \]
  \item matrose = material cost (roofing and sealing)
    \[ = 917 \, 750 \, \text{F CFA} \]
  \item matmas = material cost (masonry)
    \[ = 4 \, 818 \, 275 \, \text{F CFA} \]
  \item matcajo = material cost (carpentry and joinery)
    \[ = 2 \, 373 \, 000 \, \text{F CFA} \]
  \item matplu = material cost (plumbing)
    \[ = 755 \, 000 \, \text{F CFA} \]
\end{itemize}

The application of the formula (22) gives the following result for the estimated total cost:

\[ K_{Tf} = 12 \, 000 \, 000 \pm 5 \, 000 \, 000 \, \text{F CFA} \]

If we now apply the superficial or floor area method (SFAM) for the calculation of the total construction cost of a lodging of 110 m\(^2\) total liveable area, we find the following values according to standing:

\begin{itemize}
  \item Low standing: \( C_T = 110 \, (\text{m}^2) \times 100 \, 000 \, (\text{F CFA/m}^2) = 11 \, 000 \, 000 \, \text{F CFA} \)
  \item Average standing: \( C_T = 110 \, (\text{m}^2) \times 150 \, 000 \, (\text{F CFA/m}^2) = 16 \, 500 \, 000 \, \text{F CFA} \)
\end{itemize}

Let us specify that the unit prices per square meter are provided by the Cameroon Architect Order [25]. We note that the values obtained with SFA-Method are included in the interval envisaged by our method of estimated assessment. Let us note that the margin reached nearly 40% of the estimated total cost. That could be explained by the disparity in standing of the lodgings composing the sample having been used to validate the model. We take again the result found in b), i.e. \( K_{Tf} = 12 \, 000 \, 000 \pm 5 \, 000 \, 000 \, \text{F CFA} \). We note that the total cost could be reduced to \( 7 \, 000 \, 000 \, \text{F CFA} \), whereas the usual method fixes the total cost at \( 11 \, 000 \, 000 \, \text{F CFA} \).

\section{3.5 Total Construction Cost Reduction Zone}

The principal objective for a self-producer will be to seek to maintain the real total construction cost at the lowest level as envisaged by the estimation.

We present below the “cost interval diagram”:

\begin{itemize}
  \item \( K_T \) is the total construction cost estimation with the new method;
  \item \( C_T \) is the estimation with the usual method;
  \item \([K_T - M; K_T + M]\) is the cost interval determined with the new estimation method.
\end{itemize}

Let us consider the diagram above (1\textsuperscript{st} case). If the real total cost is maintained into the interval \([K_T - M; C_T]\), that we designate “cost reduction zone”, we will be able to then say that the reduction of the total construction cost was effective. On the other hand if the real total cost is into the interval \([K_T; K_T + M]\), baptized “overcost zone”, and then we will have spent much more money than envisaged. The interval \([C_T; K_T]\) corresponds to the “neutral zone” where the reduction or the cost overrun is not significant. In the 2\textsuperscript{nd} case, the cost reduction zone is into the interval \([K_T - M; K_T]\); the overcost zone corresponds to the interval \([C_T; K_T + M]\) and the neutral zone with the interval \([K_T; C_T]\).

\section{4. CONCLUSION}

The estimation method that we propose, while being relied on the matrix approach, is based on historical and statistical data. In a statistical method each type of construction must have a proper calibration in order to ensure the reliability of the estimation of the concerned structure. Thus for the implementation of our model we chose residential buildings in ground floor. The final
expression of the total construction cost comprises only five independent key variables (on 13 initially) selected thanks to software SPSS. Let us note that a too high number of variables raises quickly collinearity problems [14], which would mean redundancy of regressors [23]. The verification of the proposed method shows that the estimated construction costs are close to the costs calculated with the traditional methods, in particular with the SFA-Method. The total construction cost is supplied with a tolerance, a possibility which misses in the deterministic usual methods. The self-producer can thus envisage a construction budget with limits known in advance. To be maintained inside the budgetary limits remains the goal to reach to arrive to the mastering of the total construction cost. Better, to be located in the “cost reduction zone” gets a significant advantage to the self-producer which could spend definitely less money than envisaged.

For the implementation of the model, we did not take into account the management cost because those did not appear in the files where we drew our data. The method proposed has a general range basically. Nevertheless it will be advisable to adapt it to each type of constructions, especially with regard to the regression coefficients. The necessity for having a reliable data base on the construction costs is thus a major stage to sit the robustness of this estimation construction cost method.

REFERENCES


23. L.O. Brittawani, “Regression analysis: Issues of multicollinearity too often overlooked” (2004); available on: http://www.iaca.net/resources/Articles/multicollinearity_article.pdf


<table>
<thead>
<tr>
<th>Lodging</th>
<th>Area (in acres)</th>
<th>Construction Cost (in 1984 $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.219,600 F</td>
<td>3,397,000 F</td>
</tr>
<tr>
<td>2</td>
<td>1.507,900 F</td>
<td>2,254,300 F</td>
</tr>
<tr>
<td>3</td>
<td>1.509,600 F</td>
<td>2,653,200 F</td>
</tr>
<tr>
<td>4</td>
<td>1.827,000 F</td>
<td>2,046,500 F</td>
</tr>
<tr>
<td>5</td>
<td>1.391,400 F</td>
<td>1,594,000 F</td>
</tr>
<tr>
<td>6</td>
<td>1.519,600 F</td>
<td>1,854,300 F</td>
</tr>
<tr>
<td>7</td>
<td>1.397,000 F</td>
<td>2,067,000 F</td>
</tr>
<tr>
<td>8</td>
<td>1.397,000 F</td>
<td>2,067,000 F</td>
</tr>
<tr>
<td>9</td>
<td>1.739,700 F</td>
<td>2,379,000 F</td>
</tr>
<tr>
<td>10</td>
<td>1.898,000 F</td>
<td>2,654,500 F</td>
</tr>
<tr>
<td>11</td>
<td>2.091,400 F</td>
<td>2,254,300 F</td>
</tr>
<tr>
<td>12</td>
<td>2.093,700 F</td>
<td>2,379,000 F</td>
</tr>
<tr>
<td>13</td>
<td>2.338,600 F</td>
<td>2,046,500 F</td>
</tr>
<tr>
<td>14</td>
<td>2.095,000 F</td>
<td>2,254,300 F</td>
</tr>
<tr>
<td>15</td>
<td>2.343,700 F</td>
<td>2,379,000 F</td>
</tr>
<tr>
<td>16</td>
<td>2.343,700 F</td>
<td>2,379,000 F</td>
</tr>
<tr>
<td>17</td>
<td>2.343,700 F</td>
<td>2,379,000 F</td>
</tr>
<tr>
<td>18</td>
<td>2.343,700 F</td>
<td>2,379,000 F</td>
</tr>
</tbody>
</table>

Note: All costs are in 1984 dollars.

1 Euro = 63.84 F CFA