Journal of Environmental Science and Engineering A 14 (2025) 245-255
doi:10.17265/2162-5298/2025.05.004

The Sub-structure Cost Estimation Method: A Technique for Evaluating the Cost of F4-Type Buildings in Developing Countries

Alain Symphorien Ndongo[1, 2], Destin Gemetone Etou[1, 2], Westinevy Benarez Ndzessou[1, 3] and Christian Armand Anicet Tathy[1, 4]

1. Laboratory of Energetic Mechanics and Engineering (LMEI), National Polytechnic School, Marien Ngouabi University, Brazzaville BP 69, Congo

2. National Superior Polytechnic School (ENSP), University Marien Ngouabi, Brazzaville BP 69, Congo

3. Higher Institute of Architecture, Urban Planning, Building and Public Works (ISAUBTP), Dénis Sassou Nguesso University, Brazzaville BP 69, Congo

4. Higher Normal School (ENS), University Marien Ngouabi, Brazzaville BP 69, Congo

Abstract: In an environment where demand for housing is growing and the supply from public authorities is virtually non-existent, several mechanisms for housing production are emerging in the formal, semi-informal and informal construction sectors. The project owner wonders how much it costs to construct a building to an acceptable standard. Cost forecasting in general faces a number of difficulties, including a lack of available information during the preliminary phase of the project. As such, estimation becomes a crucial task involving great responsibility, which can lead to either more convincing results or chaotic situations. This study proposes a quick and effective method for estimating the cost of a single-storey F4 residential building. The modelling is done using multiple linear regression based on a statistical approach applied to twenty (20) projects that have already been completed. The project data are collected from design offices in the city of Brazzaville. The method expresses the cost of an F4 construction by certain project tasks, representing five (5) variables, three (3) of which are related to structural work and two (2) to finishing work, which are easy to determine. This approach, known as MECSO (Cost Estimation Model by Sub-structure), gives good results in all statistical tests carried out with reasonable confidence intervals. This method is very practical for engineering professionals working on the evaluation and control of construction costs.

Key words: Model, evaluation, estimation, cost, construction, F4.

1. Introduction

Housing remains a major challenge for developing countries. A World Bank study [1] highlights the difficulties in estimating the quantitative and qualitative housing deficit in emerging economies, due to a lack of harmonised data and differences in the definitions of "adequate supply". In some developing countries, it is estimated that 40% to 75% of the urban population lives in informal settlements or substandard housing, indicating a very low "adequate formal supply rate" [2]. In Africa, it is estimated that approximately 15% of the urban population can afford formal housing, given the financing costs and constraints of the formal property market [3]. According to the report "Stocktaking of the Housing Sector in Sub-Saharan Africa" [4], 85% of the urban population does not have access to formal housing loans, which significantly limits formal supply. [5]. In the Republic of Congo, the real estate sector is classified as a priority sector. According to the

Corresponding author: Destin Gemetone Etou, Ph.D., assistant professor, research field: operations research-civil engineering.

"National Development Plan (PND) for 2022-2026" [6], six sectors, namely agriculture in the broad sense, industry, special economic zones, tourism, the digital economy and real estate development, constitute the pillars for building a strong, diversified and resilient economy. Despite this prioritisation of the housing sector, demand continues to grow. Data from the *5th General Population and Housing Census of 2023 (RGPH-5)* [7] reveal that the total resident population of Congo in 2023 is 6,142,180, with an average annual growth rate of 3.2% between 2007 and 2023. The number of ordinary households stands at 1,479,197, with an average size of 4.1 persons per household. The greater the population growth, the greater the need for housing. The housing shortage is certainly due to demographic pressure in highly urbanised cities, but also to a relaxation of public authority in real estate investment programmes. New supply is insufficient to curb demand, which is rising too rapidly [8]. This area of real estate is therefore characterised by a remarkable imbalance in relation to the current rate of housing production. We are therefore seeing "self-build" projects for detached houses and even other infrastructure. In some countries, this sector accounts for nearly 90% of housing production [8]. Unfortunately, self-build projects do not often provide quality assurance for the work carried out and often lead to deadlock in the implementation of projects due to a lack of serious planning. In most developing countries in general, and in Congo in particular, we are seeing the emergence of two construction systems characterised by self-builders and semi-structured groups of companies similar to small and SMEs (Medium-Sized Enterprises) in the informal and semi-informal sectors [9]. To satisfy their legitimate need for housing, most city dwellers who do not have the necessary financial means are forced to resort to the informal and semi-informal sectors. These are the only sectors that can offer them mechanisms for producing individual housing and certain infrastructures adapted to their purchasing power and know-how. Furthermore, there are many methods of cost estimation in developing countries, but in certain aspects of project studies, there are still considerations to be exploited in order to develop others. Estimation is an assessment of the cost of a project prior to the implementation phase. Its main objective is to approximate the construction cost while remaining within limits that do not compromise the reliability of the project. Estimating the cost means being able to measure the economic weight of each component, facilitate simulation and optimisation, compare projects and adapt the justification of the amounts announced. As such, estimation becomes a crucial task involving significant responsibilities, which can lead to either more convincing results or chaotic situations. Predicting the cost of a construction project with a good degree of accuracy from the preliminary phase is beneficial for financiers, designers, contractors and project owners alike. The provisional cost estimate serves as a benchmark for measuring deviations caused by project disruptions resulting from factors such as various modifications, inflation and risks. However, cost forecasting in general faces a number of difficulties, including a lack of available information during the preliminary phase of the project. There are often significant differences between the estimated cost and the *actual* cost once the project is completed. These differences are so common with non-statistical estimation methods that it is now difficult to accurately predict the cost of a construction project using these methods. In the predominantly informal housing construction sector (more than 80%), the combination of certain factors that are poorly understood and difficult for the average builder to control, such as labour instability and the precariousness of the internal control system, compromises the effectiveness of certain common methods of estimating or controlling construction costs [10]. Cost estimation techniques based on statistical data [9, 10] have been developed considering cost as a function of materials, labour and overheads. Given the difficulties encountered in the field by some technicians, it makes sense to propose an approach based on statistical data, considering cost as a function of each trade taken independently. This is because it is more useful to estimate a project based on the set of tasks it comprises than on the set of materials, labour and overheads involved in its completion.

2. Materials and Methods

2.1 Data Collection Methods

In order to develop the statistical model, the city of Brazzaville was the geographical scope of our study. We focused exclusively on residential buildings. Focusing on the socio-economic realities of Congo, and especially those of the city of Brazzaville, and in accordance with housing standards to ensure maximum comfort, a surface area of between 80 m² and 110 m² is preferable for an average of six (6) people. This reality obviously allowed us to obtain a wide range of data on single-storey residential buildings of type F4, with average surface areas of around 100 m². The data collection operation took place in the design offices of the city of Brazzaville. We consulted sixty (60) construction files and selected twenty (20) for our study. We therefore considered the raw data from the sample of twenty (20) F4-type buildings. An F4 dwelling contains four (4) main rooms in addition to the standard kitchen, bathroom, hallway and terrace. The main rooms can be divided according to the model comprising three (3) bedrooms plus one (1) living room or lounge.

2.2 Data Analysis Method

Developing a model that takes into account all the tasks involved in the project can result in a cost model with more than ten (10) variables. In order to reduce the number of project tasks, we grouped together tasks that could be carried out either with the same materials, by the same actors, or for the same purposes. This task grouping operation allowed us to have nine (9) tasks for the development of our construction cost estimation model.

We have ten (10) variables, including one (1) variable to be explained (the project cost) and nine (9) explanatory variables corresponding to the nine (9) tasks that make up the project.

We processed the statistical data and analysed the results using Excel and statistical data processing software, namely SPSS and Statistica.

2.3 Model Construction

The method chosen to write the cost model is multiple linear regression. This allows us to write the following:

$$\mathrm{Y}_i = \boldsymbol{\alpha}_{0,i} + \boldsymbol{\alpha}_{1,i}X_{1,i} + \boldsymbol{\alpha}_{2,i}X_{2,i} + \cdots + \boldsymbol{\alpha}_{P,i}X_{P,i} + \varepsilon_i \, ; i = 1, \ldots , n \quad (1)$$

where p is the number of explanatory variables and n is the number of projects.

We thus have:

$$\mathrm{Y}_i = \boldsymbol{\alpha}_{0,i} + \boldsymbol{\alpha}_{1,i}X_{1,i} + \boldsymbol{\alpha}_{2,i}X_{2,i} + \cdots + \boldsymbol{\alpha}_{P,i}X_{9,i} + \varepsilon_i \, ; i = 1, \ldots ,20 \quad (2)$$

The various parameters defined in Table 1 are as follows:

Y: project cost;

X1: cost of preparatory work and earthworks;

X2: cost of the foundation and substructure;

X3: cost of masonry and elevation;

X4: cost of framing, roofing and plastering;

X5: cost of wood-aluminium-glass joinery and ironwork;

X6: cost of plumbing and sanitary facilities;

X7: cost of electrical and air conditioning work;

X8: cost of floor and wall coverings and painting;

X9: cost of roads and various networks.

2.4 Model Expression

Given the availability of data in the preliminary design phase, the cost estimation model would be objective if we considered variables X2 (cost of foundations and substructure), X3 (cost of masonry and elevation), X4 (cost of the roof structure, roofing and plastering), X5 (cost of wood, aluminium and glass joinery and ironwork) and X8 (cost of floor and wall coverings and paint).

Table 1 Construction costs.

No.	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
1	17,748,900	2,376,700	1,561,800	1,886,200	1,808,800	2,148,500	710,300	2,020,900	4,535,700	700,000
2	16,752,300	1,389,500	1,568,800	1,083,800	1,544,500	2,190,000	650,000	770,000	4,838,900	2,716,800
3	22,959,000	1,254,600	2,933,400	2,553,800	3,983,300	1,971,000	1,471,300	1,765,100	4,948,400	2,078,100
4	23,580,700	1,886,500	2,122,300	1,414,900	2,122,300	3,065,500	1,179,000	1,179,000	6,838,300	3,772,900
5	21,843,500	2,839,700	1,965,900	2,402,800	2,184,400	2,621,200	873,700	2,402,800	5,679,300	873,700
6	20,066,400	1,404,700	2,006,600	2,207,300	3,612,000	2,608,600	1,505,000	1,304,300	2,809,300	2,608,600
7	17,655,100	2,383,500	1,500,700	1,853,800	1,853,800	2,118,600	706,200	2,030,300	4,502,000	706,200
8	16,738,700	1,171,700	1,673,900	1,841,300	2,510,800	2,176,000	1,339,100	2,008,600	2,343,400	1,673,900
9	16,562,200	828,100	2,153,100	1,821,800	2,815,600	1,490,600	1,242,200	1,407,800	3,312,400	1,490,600
10	15,183,400	1,366,500	1,670,200	1,518,400	1,670,200	1,822,000	1,214,700	1,062,800	3,036,600	1,822,000
11	14,425,500	1,226,200	1,370,500	937,700	1,370,400	1,947,400	504,900	649,100	4,111,200	2,308,100
12	14,402,900	864,200	1,296,300	1,440,300	2,016,400	2,304,400	864,200	1,728,300	1,872,400	2,016,400
13	14,306,500	1,001,500	1,716,800	1,573,700	2,146,000	1,430,600	1,144,500	1,001,500	2,861,300	1,430,600
14	13,466,200	1,481,300	1,447,600	1,380,300	1,414,000	1,548,600	606,000	740,600	3,231,900	1,615,900
15	18,211,900	1,001,700	2,276,700	2,094,400	3,187,200	1,548,100	1,183,000	1,366,000	3,915,700	1,639,100
16	16,272,000	2,115,400	1,464,500	1,789,900	1,627,200	1,952,600	813,600	1,789,900	3,905,300	813,600
17	19,646,900	1,768,300	2,161,200	1,964,700	2,161,100	2,357,600	1,571,700	1,375,300	3,929,400	2,357,600
18	18,220,000	2,459,700	1,548,700	1,913,100	1,913,100	2,186,400	728,800	2,095,300	4,646,100	728,800
19	19,504,300	975,200	2,535,600	2,145,500	3,315,700	1,755,400	1,462,800	1,657,900	3,900,800	1,755,400
20	15,447,400	1,313,000	1,467,500	1,004,100	1,467,500	2,085,400	540,700	695,100	4,402,500	2,471,600

In our work, the model used to study the factors explaining the costs of a house is based on the construction cost function, which depends on five (5) variables as follows:

$$\boldsymbol{C = f(FS, ME, CCP, MBAVF, RSMP)} \quad (3)$$

The linear form gives us:

$$\boldsymbol{C = \alpha_0 + \alpha_1 * FS + \alpha_2 * ME + \alpha_3 * CCP + \alpha_4 * MBAVF + \alpha_5 * RSMP + \varepsilon} \quad (4)$$

where

C is the cost of the house;

FS is the amount allocated to the foundation and substructure;

ME is the amount allocated to masonry and elevation;

CCP is the amount allocated to the framework, roofing and plastering;

MBAVF is the amount allocated to wood-aluminium-glass joinery and ironwork;

RSMP is the amount allocated to floor/wall covering and painting;

$\boldsymbol{\alpha_i}$ are parameters to be estimated.

The method used to estimate the model parameters is Ordinary Least Squares (OLS).

3. Results and Discussions

3.1 Estimation Model Results

In Table 2, the dependent variable is C (construction cost). There are nine explanatory variables: TPT (cost of preparatory work and earthworks), FS (cost of foundations and substructure), ME (cost of masonry and elevation), CCP (cost of framework, roofing and plastering), MBAVF (cost of wood-aluminium-glass joinery and ironwork), PS (cost of plumbing), EC (cost of electricity and air conditioning), RSMP (cost of floor and wall coverings and painting) and VRDA (cost of roads, utilities and sanitation).

3.1.1 Data Description

Table 3 includes a description of the variables we used in our building cost estimation model. Our sample includes twenty (20) complete files for F4-type buildings. Thus, for each construction cost variable, we obtained 20 observations.

Table 2 Legend of variables.

Code	Variable	Name
Y	C	Construction cost
X1	TPT	Cost of preparatory work and earthworks
X2	FS	Cost of foundations and substructure
X3	ME	Cost of masonry and elevation
X4	CCP	Cost of framework, roofing and plastering
X5	MBAVF	Cost of wood-aluminium-glass joinery and ironwork
X6	PS	Cost of sanitary plumbing
X7	EC	Cost of electricity and air conditioning
X8	RSMP	Cost of floor and wall coverings and paint
X9	VRDA	Cost of roads, utilities and sanitation

Table 3 Description of variables.

Descriptive statistics							
	N	Range	Minimum	Maximum	Mean	Standard deviation	Variance
C	20	10,114,500	13,466,200	23,580,700	17,649,689.999	2,903,155.372	8.428×10^{12}
TPT	20	2,011,600	828,100	2,839,700	1,555,400	595,995.832	3.552×10^{11}
FS	20	1,637,100	1,296,300	2,933,400	1,822,105	432,890.192	1.873×10^{11}
ME	20	1,616,100	937,700	2,553,800	1,741,390	444,950.916	1.979×10^{11}
CCP	20	2,612,900	1,370,400	3,983,300	2,236,214.999	765,137.123	5.854×10^{11}
MBAVF	20	1,634,900	1,430,600	3,065,500	2,066,425	414,103.054	1.714×10^{11}
PS	20	1,066,800	504,900	1,571,700	1,015,585	352,899.294	1.245×10^{11}
EC	20	1,753,700	649,100	2,402,800	1,452,530	526,729.721	2.774×10^{11}
RSMP	20	4,965,900	1,872,400	6,838,300	3,981,045	1,160,989.612	1.347×10^{12}
VRDA	20	3,072,900	700,000	3,772,900	1,778,995	801,885.611	6.430×10^{11}

The elements that have the greatest weight in the expression of construction costs are those with high averages, which we found to be, in descending order: RSMP, CCP, MBAVF, FS, VRDA, ME, TPT, EC, PS.

For each variable, we obtained a range that is simply the difference between the maximum and minimum values for the variable in question. This difference is most pronounced for the RSMP and VRDA variables, which reach a total of 4,965,900 CFA francs.

The smallest standard deviation values for each category of variables are 414,103.054 CFA francs for MBAVF and 352,899.294 CFA francs for PS, respectively. This means that in our sample, the buildings are more similar in terms of MBAVF and PS costs than in terms of other variables. In other words, the observations relating to the explanatory variables MBAVF and PS are very similar compared to the other variables. Thus, the scatter plots are much more concentrated around the MBAVF and PS variables.

Variance is simply the standard deviation squared, which measures the variability that exists in a variable around the mean. The higher the variance, the more information there is to study. We note that the variables with the greatest variance are the variable to be explained C ($\mathbf{8.428} \times \mathbf{10^{12}}$) and the explanatory variable RSMP ($\mathbf{1.347} \times \mathbf{10^{12}}$). This could be explained by the disparity in the standard of the buildings in the sample used to implement the model.

3.2 Construction of the Building Cost Estimation Model

Given the unavailability of certain data during the preliminary study phase on the one hand, and the limited involvement of the various trades on the other, we decided to construct our estimation model with five explanatory variables. The variables selected are as follows: FS, ME, CCP, MBAVF and RSMP.

The linear form gives us:

$$\boldsymbol{C = \alpha_0 + \alpha_1 * FS + \alpha_2 * ME + \alpha_3 * CCP + \alpha_4 * MBAVF + \alpha_5 * RSMP + \varepsilon} \quad (5)$$

where

C is the cost of the house;

FS is the amount allocated to the foundation and substructure;

ME is the amount allocated to masonry and elevation;

CCP is the amount allocated to the framework, roofing and plastering;

MBAVF is the amount allocated to wood-aluminium-glass joinery and ironwork;

RSMP is the amount allocated to floor/wall covering and painting;

$\boldsymbol{\alpha_i}$ are parameters to be estimated.

The results obtained show us that the model is acceptable, especially since the P-value of Fisher's test is zero. There is at least one variable in the model that can explain the costs of a house.

Fig. 1 shows that all the parameters of the explanatory variables are significantly different from zero. It is easy to see that 99.67% of the cost of a house is explained by the five exogenous variables in the model, namely: FS, ME, CCP, MBAVF and RSMP. All these variables have a positive impact on the cost of a four-bedroom house. Those with the greatest impact are (i) the amount allocated to wood-aluminium-glass joinery and ironwork, and (ii) the amount allocated to the foundation and substructure. The model results show that a 1% increase in the MBAVF or FS amounts would increase the cost of the house by 2.99% or 2.42% respectively, all other things being equal. On the other hand, the variable that has

Table 4 Variables introduced (SPSS software).

Variables introduced/eliminated[a]			
Model	Variables introduced	Variables removed	Method
1	RSMP, ME, MBAVF, FS, CCP[b]	-	Introduce

a. Dependent variable: C; b. All requested variables have been introduced.

Dependent Variable: COUT
Method: Least Squares
Date: 09/27/22 Time: 16:55
Sample: 1 20
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FS	2.420364	0.289965	8.347101	0.0000
ME	1.614065	0.157028	10.27886	0.0000
CCP	0.486683	0.171073	2.844885	0.0130
MBAVF	2.992642	0.135870	22.02575	0.0000
RSMP	0.884467	0.064523	13.70781	0.0000
C	-364685.6	313673.3	-1.162629	0.2644

R-squared	0.996736	Mean dependent var	17649690
Adjusted R-squared	0.995570	S.D. dependent var	2903155.
S.E. of regression	193225.0	Akaike info criterion	27.42442
Sum squared resid	5.23E+11	Schwarz criterion	27.72314
Log likelihood	-268.2442	Hannan-Quinn criter.	27.48274
F-statistic	855.0223	Durbin-Watson stat	2.183846
Prob(F-statistic)	0.000000		

Fig. 1 Model estimation (using Statistica software).

Table 5 Model variance analysis table (SPSS software).

ANOVA[a]						
Model		Sum of squares	ddl	Mean square	F	Sig.
	Regression	159,615,208,377,076.750	5	31,923,041,675,415.350	855.022	0.000[b]
1	From student	522,702,820,923.274	14	37,335,915,780.234		
	Total	160,137,911,198,000.030	19			

a. Dependent variable: CC; b. Predictors: (Constant), RSMP, ME, MBAVF, FS, CCP.

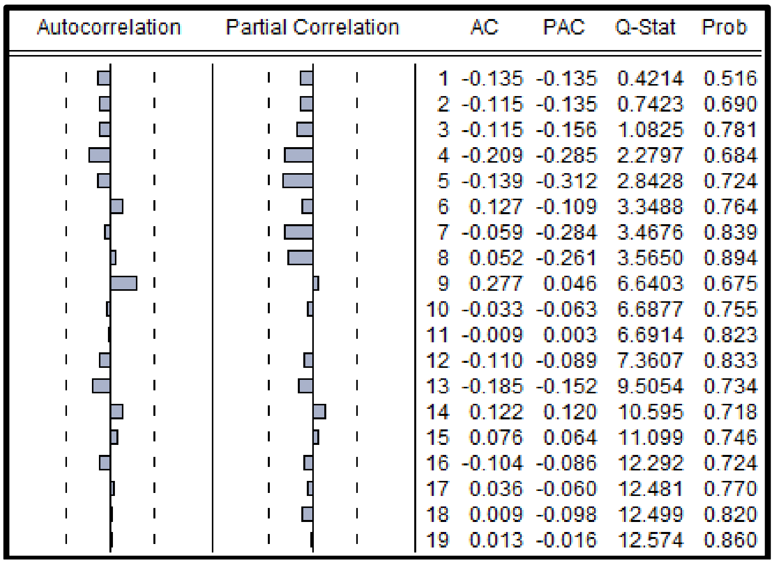

Fig. 2 Correlogram of model residuals.

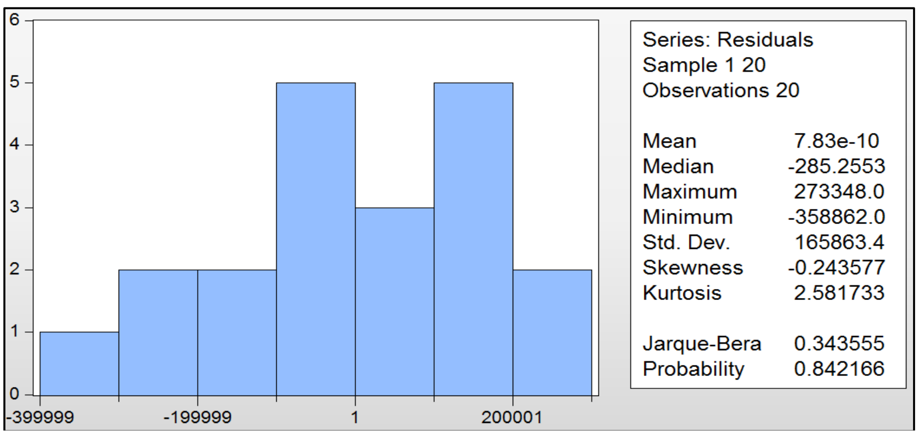

Fig. 3 Test of normality of model residuals.

the least impact on the cost of a house is the amount allocated to carpentry, roofing and plastering (CCP). The model results show that a 1% increase in CCP would increase the cost of the house by only 0.49%, all other things being equal.

3.3 Application of Statistical Tests to the Model

ANOVA is a statistical test used to verify whether several samples come from the same population or not. In our case, the analysis of variance will allow us to determine whether the observations we have obtained all come from the same family of buildings. Table 5 shows that the model obtained is significant as a whole, since *sig.* = 0.000. The first-type error risk or significance (*sig.*) provides information on the risk of misinterpreting the meaning of the regression. If *sig.* <

0.05, we can conclude that a linear regression model exists, at the 0.05 threshold (the significance threshold indicated by the *sig.* statistic).

Before concluding on the conformity of the model, we will first check whether the hypotheses made about the model are validated. These include stochastic tests (stationarity, non-autocorrelation, homoscedasticity, and normality of residuals), specification tests (Ramsey test), and model stability tests (CUSUM and CUSUM squared).

3.3.1 Stochastic Tests

The results in Fig. 2 show that the model residuals are stationary. Similarly, the results in Fig. 3 show that the P-value (0.842166) is well above 5%. This means that the residuals follow a normal distribution. The test used to study the normality of the residuals is the Jarque-Bera test.

The test used to verify the homoscedasticity of the residuals is the Breush-Pagan-Godfrey test. The observations made in Fig. 4 show that the model residuals are homoscedastic, as all P-values are greater than 5%.

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	1.619991	Prob. F(5,14)	0.2188
Obs*R-squared	7.330290	Prob. Chi-Square(5)	0.1972
Scaled explained SS	2.840667	Prob. Chi-Square(5)	0.7245

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 09/27/22 Time: 17:13
Sample: 1 20
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.88E+10	5.08E+10	-1.158075	0.2662
FS	-11638.05	46923.70	-0.248021	0.8077
ME	20061.61	25411.09	0.789482	0.4430
CCP	-9459.626	27683.97	-0.341700	0.7377
MBAVF	40594.92	21987.28	1.846291	0.0861
RSMP	2124.455	10441.46	0.203463	0.8417

R-squared	0.366515	Mean dependent var	2.61E+10
Adjusted R-squared	0.140270	S.D. dependent var	3.37E+10
S.E. of regression	3.13E+10	Akaike info criterion	51.41297
Sum squared resid	1.37E+22	Schwarz criterion	51.71169
Log likelihood	-508.1297	Hannan-Quinn criter.	51.47129
F-statistic	1.619991	Durbin-Watson stat	1.700759
Prob(F-statistic)	0.218842		

Fig. 4 Test of homoscedasticity of the model residuals.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.318100	Prob. F(2,12)	0.7335
Obs*R-squared	1.006949	Prob. Chi-Square(2)	0.6044

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 09/27/22 Time: 17:15
Sample: 1 20
Included observations: 20
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FS	0.105968	0.337139	0.314315	0.7587
ME	0.014656	0.191308	0.076611	0.9402
CCP	-0.057270	0.213237	-0.268575	0.7928
MBAVF	0.004676	0.152153	0.030733	0.9760
RSMP	-0.015101	0.071213	-0.212053	0.8356
C	-35514.63	333162.8	-0.106598	0.9169
RESID(-1)	-0.232387	0.374838	-0.619966	0.5469
RESID(-2)	-0.165833	0.346872	-0.478082	0.6412

R-squared	0.050347	Mean dependent var	7.83E-10
Adjusted R-squared	-0.503617	S.D. dependent var	165863.4
S.E. of regression	203385.1	Akaike info criterion	27.57276
Sum squared resid	4.96E+11	Schwarz criterion	27.97106
Log likelihood	-267.7276	Hannan-Quinn criter.	27.65052
F-statistic	0.090886	Durbin-Watson stat	1.989274
Prob(F-statistic)	0.997994		

Fig. 5 Test for non-autocorrelation of model residuals.

To test for non-autocorrelation of the residuals, the most commonly used test is the Breusch-Godfrey test. This test is based on Fisher's null hypothesis test for coefficients (F-statistic) or the Lagrange Multiplier (LM-test), whose test statistic is nR^2. The objective of the test is to find a significant relationship between the residual and the same residual shifted. Fig. 5 shows the results of the Breusch-Godfrey test for $p = 2$ (when $p = 1$, we refer to the Durbin-Watson (DW) test). Looking at this figure, we see that no lag is significant (all P-values for RESID (-1) and RESID (-2) are greater than 5%). We therefore conclude that there is no autocorrelation of errors.

3.3.2 Ramsey Tests

This test allows us to determine whether we have forgotten to include a very important variable in the model. To this end, the results in Fig. 5 reveal that no important variables have been omitted from the model. Indeed, as the P-value of the variable "FITTED^2" is greater than 5%, we can conclude that the model is well specified.

3.3.3 Model Stability Tests

To ensure the overall stability of the model parameters, CUSUM and CUSUM squared tests are often used. The latter allows structural instabilities to be detected. Looking at the graphs in Fig. 7, we can conclude that the model parameters are structurally stable (based on the results of the first graph) and pointwise stable (based on the results of the second graph). This is because the curves in both graphs are entirely contained within the area formed by the straight lines representing the confidence threshold.

3.4 Expression of the Cost Estimation Model

$$\boldsymbol{C = -364,685.6 + 2.42 * FS + 1.61 * ME + 0.49 * CCP + 2.99 * MBAVF + 0.88 * RSMP \pm M} \quad (6)$$

with

$\alpha_0 = -364{,}685.6$; $\alpha_1 = 2.42$; $\alpha_2 = 1.61$; $\alpha_3 = 0.49$; $\alpha_4 = 2.99$; $\alpha_5 = 0.88$.

These regression coefficients are given with a 95.0% confidence interval.

α_0 : $[-1{,}037{,}447.933 ; 308{,}076.735]$;
α_1 : $[1.798 ; 3.042]$;
α_2 : $[1.277 ; 1.950]$;
α_3 : $[0.119 ; 0.853]$;
α_4 : $[2.701 ; 3.284]$;
α_5 : $[0.746 ; 1.022]$.

For a simple and clear presentation of the MECSO (Cost Estimation Model by Sub-structure), we use the notation CSO_i for the independent variables selected in the final expression of the model in order to facilitate user understanding. Thus, we have:

$$\boldsymbol{C = \alpha_0 + \alpha_1 * CSO_1 + \alpha_2 * CSO_2 + \alpha_3 * CSO_3 + \alpha_4 * CSO_4 + \alpha_5 * CSO_5 \pm M} \quad (7)$$

Ramsey RESET Test
Equation: EQUATION1
Specification: COUT FS ME CCP MBAVF RSMP C
Omitted Variables: Squares of fitted values

	Value	df	Probability
t-statistic	0.830268	13	0.4214
F-statistic	0.689345	(1, 13)	0.4214
Likelihood ratio	1.033369	1	0.3094

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	2.63E+10	1	2.63E+10
Restricted SSR	5.23E+11	14	3.73E+10
Unrestricted SSR	4.96E+11	13	3.82E+10

LR test summary:

	Value
Restricted LogL	-268.2442
Unrestricted LogL	-267.7276

Unrestricted Test Equation:
Dependent Variable: COUT
Method: Least Squares
Date: 09/27/22 Time: 17:20
Sample: 1 20
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FS	2.910332	0.658972	4.416471	0.0007
ME	1.864937	0.341345	5.463497	0.0001
CCP	0.554116	0.191119	2.899320	0.0124
MBAVF	3.536818	0.669670	5.281434	0.0001
RSMP	1.038706	0.196896	5.275414	0.0002
C	-2045117.	2048669.	-0.998266	0.3364
FITTED^2	-4.82E-09	5.80E-09	-0.830268	0.4214

R-squared	0.996900	Mean dependent var	17649690
Adjusted R-squared	0.995470	S.D. dependent var	2903155.
S.E. of regression	195405.2	Akaike info criterion	27.47276
Sum squared resid	4.96E+11	Schwarz criterion	27.82126
Log likelihood	-267.7276	Hannan-Quinn criter.	27.54079
F-statistic	696.8230	Durbin-Watson stat	2.213915
Prob(F-statistic)	0.000000		

Fig. 6 Ramsey test of the model.

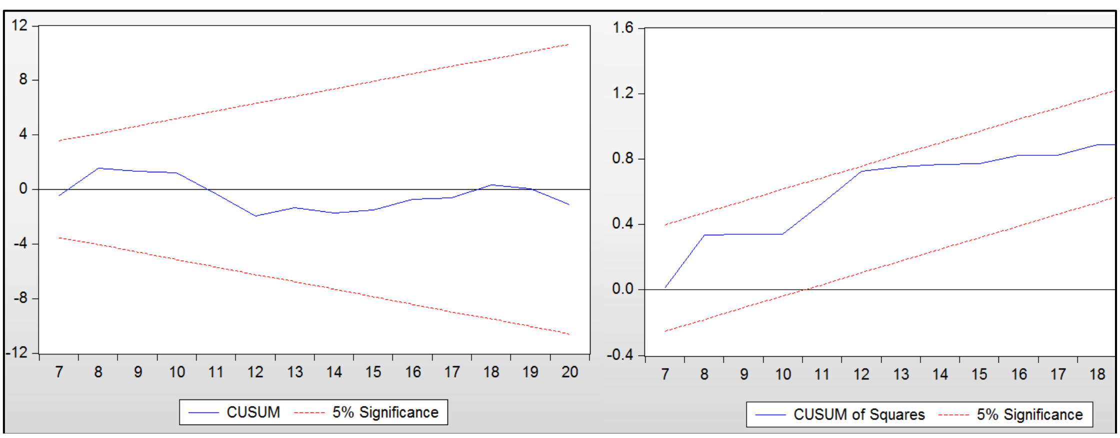

Fig. 7 Model stability test.

with

$\alpha_0 = -364{,}685.6$; $\alpha_1 = 2.42$; $\alpha_2 = 1.61$; $\alpha_3 = 0.49$; $\alpha_4 = 2.99$; $\alpha_5 = 0.88$.

$\boldsymbol{CSO_1}$: is the amount allocated to the foundation and base;

$\boldsymbol{CSO_2}$: is the amount allocated to masonry and elevation;

$\boldsymbol{CSO_3}$: is the amount allocated to the framework, roofing and plastering;

$\boldsymbol{CSO_4}$: is the amount allocated to wood-aluminium-glass joinery and ironwork;

$\boldsymbol{CSO_5}$: is the amount allocated to floor/wall covering and painting;

$\boldsymbol{M}$: is the tolerance margin.

3.4.1 Model Verification

Let us consider dwelling No. 7 in Table 1. Applying formula (4.15) gives:

C = -364,685.6 + 2.42 (1,500,700) + 2.42 (1,853,800) + 1.61 (1,853,800) + 0.49 (2,118,600) + 0.88 (4,502,000) = 17,456,362 CFA francs.

We note that the estimated total cost C is very close to the actual total cost K (17,655,100). The residual value is very low : $\boldsymbol{\varepsilon}$ = C - C' = 198,739 CFA francs.

Using the same method for dwelling No. 20, we find a total estimated cost of 15,447,400 CFA francs, which is very close to the actual total cost (15,631,886 CFA francs), with a residual of -184,486 CFA francs.

Unlike the statistical matrix method [11], which defines cost as the sum of material, labour and overhead costs, the sub-work cost estimation method estimates the cost of certain project tasks that are easy to determine.

This method is very practical for civil engineering professionals who find it difficult to determine the costs of certain tasks that are often included in building equipment, such as plumbing, electrical work and air conditioning.

4. Conclusion

There are many estimation techniques, but few allow costs to be estimated with low margins of error. Estimation at the preliminary design stage should provide the project owner with information on the expected costs of carrying out the work and assist in the preparation and implementation of procurement procedures. The estimation method studied is a decision-making tool that can facilitate the evaluation of a project at the preliminary design stage. Based on scientific reasoning and mathematical tools dedicated to problem modelling, the resulting cost estimation model, known as MECSO (model for estimating the cost of construction by sub-structure), makes it possible to determine the construction price without calculating all the project's building trades. This makes it possible to circumvent certain difficulties in evaluating work items that require specific knowledge of the field. The cost obtained is very close to the actual cost, as the results of tests applied to the model are very satisfactory, with more or less reasonable deviations. This estimation tool remains practical in that it has been designed on a secure and reliable basis of projects already completed.

References

[1] World Bank. 2021. "Introducing the Adequate Housing Index (AHI)." documents1.worldbank.org/curated/en/370141635794493951/pdf/Introducing-the-Adequate-Housing-Index-AHI-A-New-Approach-to-Estimate-the-Adequate-Housing-Deficit-within-and-across-Emerging-Economies.pdf.

[2] Habitat for Humanity. n.d. "What Is Affordable Housing?" www.habitat.org/emea/about/what-we-do/affordable-housing.

[3] ACRC. 2024. www.african-cities.org/wp-content/uploads/2024/07/ACRC_Working-Paper-18_July-2024.pdf.

[4] World Bank. 2015. "Stocktaking of the Housing Sector in Sub-Saharan Africa: Challenges and Opportunities" https://www.worldbank.org/content/dam/Worldbank/document/Africa/Report/stocktaking-of-the-housing-sector-in-sub-saharan-africa-full-report.pdf.

[5] World Bank. 2015. "Stocktaking of the Housing Sector in Sub-Saharan Africa: Summary Report." www.worldbank.org/content/dam/Worldbank/document/Africa/Report/stocktaking-of-the-housing-sector-in-sub-saharan-africa-summary-report.pdf.

[6] Cadre Strategique de Developpement (CSD). 2022. "National Development Plan (PND) of the Republic of

Congo for the 2022-2026 Financial Years.” https://gouvernement.cg/wp-content/uploads/2022/07/CSD-PND-2022-2026.pdf.

[7] National Institute of Statistics (INS). 2023. *Fifth General Population and Housing Census (RGPH-5): Preliminary Results*. Brazzaville: INS. (in French)

[8] Destin Gemetone Etou. 2021. “The Dysfunctions of Construction Project Management in Developing Countries: The Case of the Republic of Congo.” *American Journal of Engineering Research (AJER)* 10 (12): 202-7.

[9] Louzolo-Kimbembe, P. February 2005. “Contribution to Methods of Controlling Construction Costs in Developing Countries: Cost Estimation and Optimisation in a Non-time-bound Context.” Ph.D. thesis, Ecole Nationale Supérieure de Yaoundé (Cameroon).

[10] Manjia, M. B. March 2011. “A Model for Forecasting Construction Costs in Developing Countries: The Case of Cameroon. From Multiple Linear Regression to Artificial Neural Networks.” Ph.D. thesis, Ecole Nationale Supérieure de Yaoundé (Cameroon).

[11] Louzolo-Kimbembe, P., and Pettang, C. 2006. “Contribution to the Amelioration of the Estimation Method of Construction Cost’s Mastering in Developing Countries.” *International Journal on Architectural Sciences* 7 (1): 14-25.