Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2021 doi:10.20944/preprints202110.0201.v1

COST MODELLING FROM THE CONTRACTOR PERSPECTIVE: APPLICATION TO RESIDENTIAL AND OFFICE BUILDINGS

Francisco Pereira Monteiro[a], Vitor Sousa[b], Inês Meireles[c], Carlos Oliveira Cruz[b]*

[a]Department of Civil Engineering, Architecture and GeoResources, Universidade de Lisboa – IST, Av. Rovisco Pais, 1049-001 Lisbon, Portugal

[b]CERIS, Department of Civil Engineering, Architecture and GeoResources, Universidade de Lisboa – IST, Av. Rovisco Pais, 1049-001 Lisbon, Portugal

[c]RISCO; Department of Civil Engineering, University of Aveiro, Campus de Santiago, 3810-193 Aveiro, Portugal

ABSTRACT

For the majority of the contractual arrangements used in construction projects, the owner is not responsible for the cost deviations due to the variability of labor productivity or material price, amongst many other aspects. Consequently, the cost performance of a project may be entirely distinct for the owner and the contractor. Since the majority of the quantitative research on cost estimation and deviation found in the literature adopts the owners' perspective, this research provides a contribution towards modelling costs and cost deviation from a contractors' perspective. From an initial sample of 13 residential building and 10 office building projects, it was possible to develop models for cost estimation at the early stage of development including both endogenous and exogenous variables. Although the sample is relatively small, the authors were able to fully analyze all the cost data, using no secondary sources of data (very frequent in cost modelling studies). The statistically significant variables in the cost estimation models were the areas above and below ground and the years following the 2008 financial crisis, including the international bailout (2011-2014) period. For estimating the unit cost, a nonlinear model was obtained with the number of underground and total floor, the floor ratio and the years following the 2008 financial crisis, including the international bailout (2011-2014) period as predictors. For the office buildings, it was also found a statistically significant correlation between the cost deviation and the number of underground floors.

Keywords: cost estimation, cost deviation, financial crisis, promotor-contractor, statistical modelling

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2021 doi:10.20944/preprints202110.0201.v1

1 INTRODUCTION

Construction projects complexity is increasing, both on their "hard" (or tangible) and "soft" (or intangible) dimensions. From new materials to new construction technologies, a multitude of technical solutions emerged over the last decades, widening the range of alternative options available for the "hard" dimension of construction projects. Concurrently, the range of aspects to manage in construction project has also increased. The "soft" dimension of construction projects includes the need for satisfying an increasingly broader and stringent social (e.g., health and safety), environmental (e.g., construction and demolition waste management) and economic (e.g., use of life-cycle cost as awarding criteria on public projects in the European Union) requirements. Consequently, construction managers are now facing additional challenges in their projects. To aid them in their tasks, several standards and regulations have been published (e.g. ISO 21500 family of standards) and new tools are becoming available (e.g., BIM – Building Information Modeling). These provide holistic and consistent guidelines and technological support to tackle the complexity of managing construction projects within this new context.

Despite all these evolutions, the financial control of construction projects is still a dominant dimension in the project's governance. In this regard, cost estimation at the early stages dictates the investment decisions, although, at the early stages, there is a significant risk surrounding the estimation, given the technical uncertainty. Therefore, more accurate cost forecasting at the early stages of the project's development and better quantification/understanding of cost deviations are amongst the key concerns of any construction project manager (Hegazy, 2002).

Within this research, the contractor perspective is adopted by analyzing the financial performance of 23 building projects of a large industrial group in Portugal (13 residential building and 10 office building projects). Among the companies in the group, there is real estate and a contractor that develop, amongst other types of projects, residential and office buildings in collaboration. Although the dataset is relatively small, it is homogenous, in the sense that the contractor was the same company, and the cost analysis used no secondary data. The real estate assumes all the licensing, design, marketing and commercialization and the contractor executes the projects. The contractor develops also projects for external clients, both private and public, of various types (e.g., commercial, healthcare and educational buildings; water, transportation and energy infrastructures).

The paper is organized as follows: after the introduction, section 2 presents the literature review; section 3 explains the data used and the methods; section 4 presents the results; and, finally, section 5 provides the main conclusions.

2 LITERATURE REVIEW

Historically, there have been several tolls for cost estimating at early stages of the project's development. The simplest models are based on parametric estimation of costs, built upon expert judgments (see for instance Cruz and Branco, 2020). The traditional multiple regression analysis (RA) has been the tool most used by researchers (e.g., Lowe et al., 2006; Trost and Oberlender, 2003). Artificial neural networks (ANN) have gained some expression for data modeling in various engineering problems, including cost estimation (e.g., Kim et al. 2005; Sonmez 2004; Hegazy and Ayed 1998), and case-based reasoning (CBR) is also being used in various tasks related to construction management (e.g., resource estimation – Soto et al., 2017; duration estimation – Jin et al., 2016). A review on CBR use for construction management can be found in Hu et al. (2016) and its use for cost estimation can be found in Kim and Kim (2005), Ji et al. (2010) or Ahn et al. (2020). A comparison between the three methods was done by Kim et al. (2004), with the new tools achieving better results than regression models. More recently, Rafiei and Adeli (2018) developed cost estimation models using support vector machines, along with ANN combined with an unsupervised deep Boltzmann machine, and included exogenous variables (e.g., consumer price index, interest rate for loan, population of the city) in combination with endogenous variables (e.g., total area). Some authors have also developed models to estimate the cost of portions of the projects (e.g., structure - Günaydin and Doğan, 2004; Doğan et al., 2006).

Table 1 lists summarizes the main research on the topic, along with the methods and explanatory variables used in each study. It should be noted that some models were developed to estimate the total cost (when the area is included in the model) whereas others were developed to estimate the unit cost (when the area is not included in the model). Some variables listed in Table 1 should be interpreted as a category of variables rather than a single variable, in some cases simply because they are measured differently depending on the author. For instance, the construction area may be gross, usable, or other; the number of stories may also be total, above ground and underground; the height may be of the building or of the floor. Others are naturally a category of

 doi:10.20944/preprints202110.0201.v1

variables, such as the structural characteristics that may include the type of structure or foundation (e.g., Jin et al., 2012). A few are even impossible to quantify adequately at the early stages of the project development, namely the duration. In fact, it is far more common to use cost as an independent variable to estimate the construction duration (e.g., see Sousa et al. (2014a,b) or Sousa and Meireles (2018) for examples of time-relationships), because cost estimate tends to be done by the designer before the contractor develops the construction schedule.

There are also authors attempting to use BIM for conceptual cost estimation (e.g., Muratova and Ptukhina, 2019). However, this approach requires a quantities takeoff, which implies a degree of project development that is incompatible with the early stages of development in this research (definition of general characteristics of the project, such as area and number of floors, and a preliminary sketch). In fact, even some models reported in the literature review presented herein use variables that may be unavailable at this stage of the project development (e.g., proportion of walls and windows in the external envelope). There is a clear trade-off between model adjustment, i.e., estimation accuracy and the availability of information in the early stages of the project. The review presented was focused on cost estimation for building projects and it is not intended to be exhaustive, but rather illustrate that different tools, sample sizes and variables have been used. There is also an extensive literature on other types of projects (e.g., transportation infrastructure projects - Karaca et al. 2020; Swei et al., 2017; Flyvbjerg et al., 2016; Gunduz et al., 2011; Al-Tabtabai et al. 1999).

The topic of cost deviations is closely related to cost estimation, since a more accurate cost estimation should reduce cost deviations. There is an extensive literature on the magnitude (e.g., Shehu et al., 2014; Sweis et al., 2013; Love et al., 2013, 2015, 2019, 2020) and causes (e.g., Kaming et al., 1997; Abusafiya and Suliman, 2017; Derakhshanalavijeh and Teixeira, 2017; Annamalaisami and Kuppuswamy, 2019; Balali et al., 2020) of cost deviations. The former tends to be quantitative, based on the analysis of the performance of past project, while the latter is mostly qualitative, resorting to questionnaires or interviews with experts.

Table 1 – Early-stage cost estimation models for buildings

Reference	Thalmann (1998)	Elhag and Boussabaine (1998)	Emsley et al. (2002)	Picken and Ilozor (2003)	Attalla and Hegazy (2003)	Skitmore and Ng (2003)	Elhag et al. (2005)	Li et al. (2005)	Stoy and Schalcher (2007)	Wheaton and Simonton (2007)	Stoy et al. (2008)	Jin et al. (2014)	Bayram et al. (2016)
Method													
RA	X		X	X	X	X		X	X	X	X		
ANN		X	X		X								X
CBR												X	
Other							X						
Variables													
Project related													
Building type		X				X		X		X			
Area	X	X	X					X		X	X	X	X
Number of stories		X	X							X		X	X
Number of households												X	
Height			X	X				X	X				X
Duration		X	X		X				X		X		
Location			X								X		
Above ground external envelope characteristics	X		X								X		
Underground external envelope characteristics	X		X										
Number of lifts			X								X	X	
Number of piloti floors												X	
Structural characteristics			X										
Other			X		X		X		X		X	X	
Management related													
Type of contract			X			X	X						
Procurement strategy			X		X	X	X						
Other			X		X								
Other													
Type of client						X	X						
Construction year	X									X			
Designer characteristics							X						
Contractor characteristics								X					
Site characteristics			X										
Sample													
Size	15	30	288	36	50	93	-	30	290	42340 18469	75	91	232
Type	R	S	R				-	O		R O	R	R	

R – Residential buildings

O – Office buildings

S – School buildings

The research relating the magnitude with the causes of cost deviation is less extensive and the causes are limited to macro variables of the projects, such as: i) the size of the project (Shrestha and Fathi, 2019; Flyvbjerg et al., 2004); ii) the nature of ownership/promotor (public or private – e.g., Flyvbjerg et al., 2002; Sweis et al., 2013);

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2021 doi:10.20944/preprints202110.0201.v1

iii) the type of intervention (new build or refurbishment/rehabilitation – Shehu et al., 2014); iv) the type of project (residential, infrastructure, commercial, and other – e.g., Pearl et al. 2003); v) the procurement model (design-bid-build, design and build, project management - e.g., Buccciol et al., 2013; Shrestha and Fathi, 2019); or vi) the tender method (open, selection, negotiated tendering – e.g., Reyers and Mansfield, 2001).

Most research on cost modeling in general (cost estimation and cost deviations) tends to focus on variables endogenous to the projects. Table 1 is provides a clear illustration of this claim, with the variables used by the various authors being exclusively related to the project or its management. There is a smaller body of literature on the influence of exogenous variables on the financial performance of construction projects. For instance, Catalão et al. (2019a,b, 2020) demonstrated the relation between political and economic cycles and the cost deviation in public projects.

The quantitative research available in the literature, both in terms of cost estimation and quantitative analysis of cost variations, tends to reflect the construction projects' financial performance from the owners' perspective. The records used by most of the authors were obtained from the owners (or from the contractors) and represent the payments made to the contractors and not the expenses of the contractors. However, the amounts payed by the owners do not match perfectly the amounts spent by the contractors to execute the projects after deducting the profit margin. Regarding the cost estimation, the owners' perspective is affected by the commercial strategy adopted by the contractors in each moment, frequently represented by the margin defined in their bids. In highly competitive contexts, the margins tend to decrease, whereas in low competitive contexts the margins tend to increase. Concerning cost deviations, the variability of materials prices, labor productivity or site overheads, amongst other potential causes of cost deviation (e.g., accidents, equipment breakdown or failure) are not measured when analyzing historical construction cost data from the owners' perspective. From the owners' perspective, change orders and errors/omissions (if the design is provided by the owner) are the most relevant causes of cost deviations.

The literature has provided recently an active discussion whether cost deviations are motivated by more technical aspects (e.g., cost escalation, scope changes, unforeseen events/conditions) of the projects (Love and Ahiaga-Dagbui, 2018; Love et al. 2019) or by estimator bias (Flyvbjerg et al., 2018, 2019). However, this discussion is outside of the scope of the present research. This discussion is focusing on the cost deviations between the first estimate and the final cost, and in the context of major infrastructure projects more applicable to public projects. This includes references to the benefits of the projects for the society. Herein, the scope is restricted to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2021 doi:10.20944/preprints202110.0201.v1

private projects and cost deviations between the detailed design and final cost. Furthermore, the cost-benefit ratio is simply the cost of the project versus the income generated by its commercialization. So, fundamentally the technical aspects will drive the cost deviations and the potential estimator bias will be more on the expected market valuation of the project.

3 DATA AND METHODS

As referred above, the data used was obtained from a large industrial group in Portugal that include a real estate and a contractor in their portfolio of companies. All projects were developed in collaboration between these two companies of the group and, despite the formal split between then, they end up working as single entity with complementary expertise.

The 23 building projects were developed mostly in Portugal, with only 2 being abroad (Angola and Mozambique). The projects in Portugal are concentrated in the Lisbon and Porto metropolitan areas (the two major cities in Portugal) and can be classified as premium. The information on the projects includes the: i) proportion of the cost by major category of works (structure, architecture, technical installations and site overheads); ii) estimated cost; iii) profit margin; iv) estimated price; v) final price; vi) total area, above ground and underground gross-built area; and vii) total floors, above ground and underground number of floors. There is also information on the start year and duration of the projects. Both the cost and prices of the projects were update to 2019 values using the formulas for price adjustment applicable to public residential and office buildings in Portugal. In Portugal, the reimbursements to contractors in public construction projects are corrected to account for inflation. Since this is mandatory, there are formulas defined by law for estimating the increase (or decrease) in the payments to the contractor for 23 different types of projects (Law-Decree nº 6/2004). These formulas represent the average weight of labor, materials (a selection from 51 different materials) and equipment on the total price of the projects. The price indexes of the labor, materials and equipment are published monthly by the government based on the official inflation data. The estimated and final unit prices and the cost deviations were calculated from the available data. Not all fields were possible to retrieve for all the projects, particularly the final price that was available for only 16 projects.

In addition to the endogenous variables, the influence of the 2008 financial crisis and subsequent international bailout that Portugal had between 2011 and 2014 was also included. This exogenous variable was modeled with a categorical predictor assuming the value of 1 between 2008 and 2014 and 0 in the remaining years. A lag of 1 year was also considered at the start and end of the crisis to evaluate if there was a delay between these events and the impact on the cost of the projects.

Due to confidentiality issues regarding some of the data (revealing the cost without the profit margin of the contractor for an external client), indexes were computed dividing the value of each project by the average of all the projects in the sample. This was done particularly for the projects profit margin, total and unit cost, and total and unit initial and final prices. Area and floor ratios were also computed dividing the values above ground by the values underground since there is typically a relation both due to parking requirements.

A statistical approach was used to analyze the data, comprising of two steps: i) a preliminary data analysis; and ii) a data modeling. The preliminary data analysis included calculation of descriptive statistics, assumptions testing and unidimensional statistical analysis. The normality and homogeneity of variance were tested using the Shapiro-Wilk and Levene's tests, respectively, and the unidimensional analysis was done using either parametric or non-parametric distribution comparison (t-test / ANOVA or Mann-Whitney/Kruskal-Wallis), for categorical variables, and correlation (Pearson or Spearman), for continuous variables. The data modeling was based on the traditional least squares multiple linear regression. Non-linear regression was also used, when necessary, but given the sample size the use of artificial intelligence tools (e.g., artificial neural networks, support vector machines, random forests) was not considered. Given the small sample size, bootstrapping (1000 simulations with simple sampling and 95% confidence interval based on percentile) was used to strengthen the confidence in the results.

The restriction of the context (projects from a single company), scope (all buildings are classified as premium in terms of quality) and location (the spatial variability of the locations is small) limits the generalization of the results. However, it excludes these variables from the cost estimation and deviations of the projects and enables the possibility of capturing the cost estimation and deviations drivers that are specific to the projects. This is an important difference from most past research effort, which in most cases use data samples with projects that may be very different, developed by distinct contractors, designed by different teams and, in some cases, promoted by various owners in many locations. This broader scope allows capturing an overall average cost performance of the projects, but it is impossible to assess if it was due to the contractor competence, design quality, owner experience, nature of the project, local factors or other aspects that are controlled for in the analysis. Consequently, using large mixed samples of data may fail in terms of applicability to a specific project.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2021 doi:10.20944/preprints202110.0201.v1

4 RESULTS AND DISCUSSION

The projects totalize a cost of over 155 million euros, with the residential buildings contributing with 57% and the office buildings 43%. The initial price (cost plus typical margin used by the contractor for external clients) of each individual project ranged from 1.5 to 20 million euros. The average initial unit price is 560 €/m^2, for office buildings, and 785 €/m^2, for residential buildings. This difference is, however, strongly influenced by the two residential buildings outside Portugal (one in Angola and another in Mozambique) that had an average initial unit price of 1 408 €/m^2. Table 2 presents some descriptive statistics characterizing the dataset.

Figure 1 illustrates the distribution of the cost and price indexes and the weight of each cost category for the residential and office buildings. The number of projects with information regarding the cost and initial price is roughly the same, but there are fewer projects with information regarding the final price. Consequently, analyzing the evolution from cost to final price is not possible (Figure 1 bottom). Considering the substantial price difference of the projects outside Portugal, one of them clearly an outlier identified in Figure 1, they were excluded from the analysis from this point forward.

Comparing the weight of the cost categories between residential and office building, it is visible a difference in all cost categories except for the site overheads. These differences were found to be statistically significant (Table 3), and the site overheads would also be considered statistically significant for a significance level of 0.10 instead of the typical 0.05. The parametric t-test was used since the data was found to be normally distributed for both residential and office buildings subsets according with the Shapiro-Wilk test.

Table 2 – Descriptive statistics of some of the main variables in the dataset

Variable		Sample	Range	Minimum	Maximum	Sum	Mean	Std. Dev.	Skewness	Kurtosis
Floors [-]	Underground	19	4	1	5	64	3.37	1.065	-0.849	1.152
	Above Ground	21	20	3	23	153	7.29	4.880	2.162	4.987
	Total	19	16	5	21	189	9.95	3.837	1.424	2.528
	Ratio	19	6.25	0.75	7.00	42.48	2.24	1.531	1.944	4.278
Area [m^2]	Underground	22	16 893.00	420.00	17 313.00	131 353.75	5 970.63	4 184.18	0.905	0.926
	Above Ground	22	10 342.00	1 557.00	11 899.00	142 095.44	6 458.88	2 983.97	0.221	-0.740
	Total	23	26 136.00	1 977.00	28 113.00	294 621.19	12 809.62	6 671.70	0.287	-0.311
	Ratio	22	3.08	0.62	3.71	33.14	1.51	0.833	0.935	0.661
Cost Category Weight [%]	Structure	23	20.00	12.70	32.70	540.30	23.49	5.019	-0.237	-0.398
	Architecture	23	25.10	29.60	54.70	955.70	41.55	7.751	-0.128	-1.420
	Technical Installations	23	24.90	9.50	34.40	532.60	23.16	6.822	-0.425	-0.449
	Site Overheads	23	12.50	7.50	20.00	263.40	11.45	2.988	1.598	2.854
Total Cost Index [-]		21	21	2.11	0.19	2.29	21.00	0.126	0.333	0.501
Margin Index [-]		21	21	1.96	0.36	2.32	21.00	0.134	0.375	0.501
Price [-]	Initial	22	19 367 364.57	1 477 203.03	20 844 567.61	185 850 166.26	8 447 734.83	5 107 220.52	0.873	0.610
	Final	16	19 159 444.23	2 746 435.50	21 905 879.73	155 809 085.39	9 738 067.84	5 404 338.41	0.970	0.421
Unit Price [-]	Initial	22	1 401.44	429.25	1 830.69	15 022.83	682.86	288.56	3.239	12.577
	Final	16	1 441.43	402.96	1 844.39	11 576.50	723.53	343.78	2.563	7.779
Cost Deviation [%]		15	15	38.06	-13.41	24.66	57.00	2.153	69.507	0.580
Duration [days]		23	23	240	240	480	7320	14.109	4578.656	0.481

Note: the margin and cost, both total and unit values, were not included due to confidentiality of the data

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2021 doi:10.20944/preprints202110.0201.v1

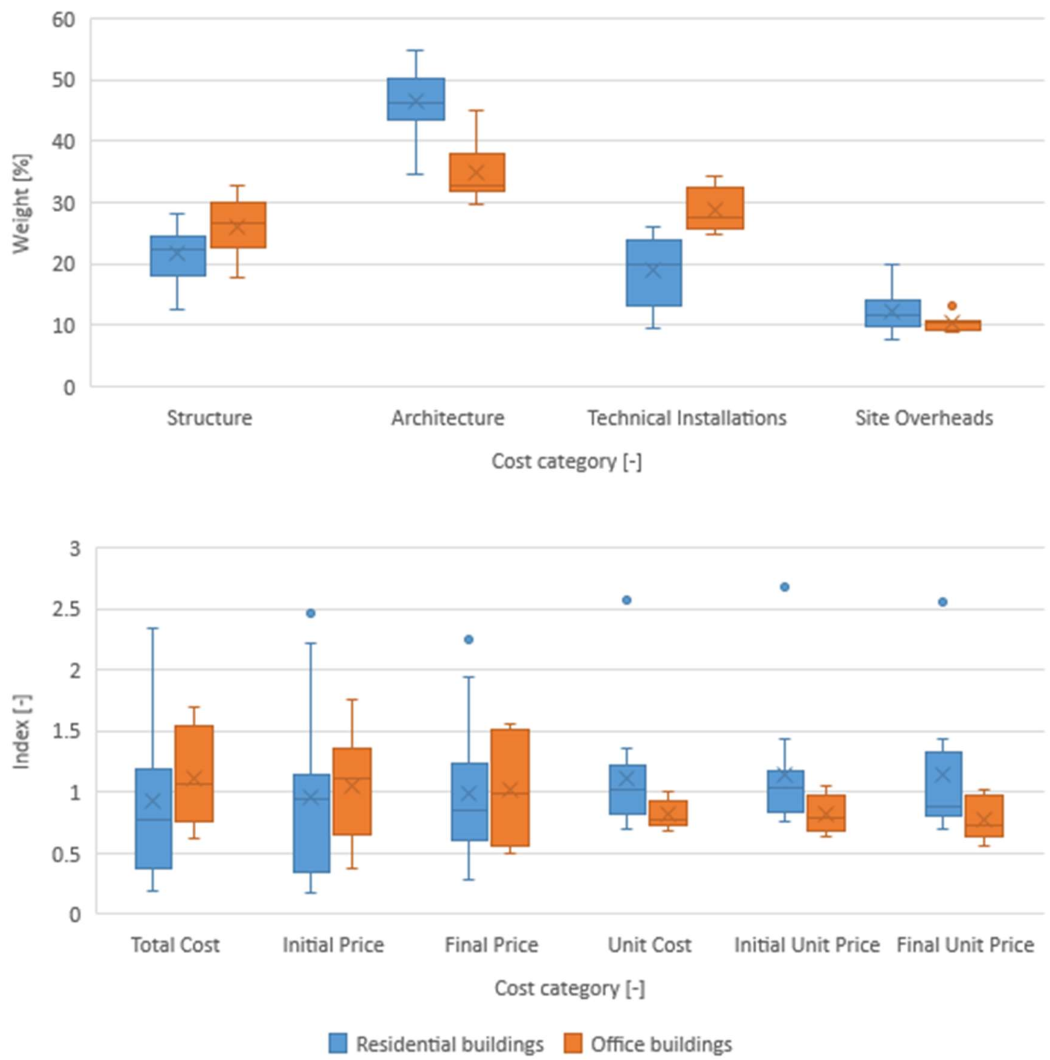

Figure 1 – Projects distribution of the weight by cost category (top) and cost and prices indexes (bottom)

Office buildings present a lower weight of architecture costs, which may be explained by the tendency for open spaces. These savings are partially compensated by more expensive structures and technical installations, since the unit cost difference is only statistically significant for a 10% significance level. Assuming that the open spaces imply wider spans, this may contribute to explain the higher weight of the structures in office buildings. The demand for heating, ventilation and air conditioning and the requirements regarding the electric and telecommunication facilities tend to be higher for office buildings than for residential buildings, which may explain the results. These results were further confirmed by bootstrapping (not presented herein the full table of results), with the significance of the t-test result increasing to 0.045, 0.003 and 0.002, for the structure, architecture, and technical installations, respectively.

 doi:10.20944/preprints202110.0201.v1

Table 3 – Means comparison between residential and office buildings cost categories weights and unit cost and prices

Variables		Levene's Test		t-test			Difference			
		F	Sig.	t	Df	Sig. (2-tailed)	Mean	Std. Error	95% Confidence Interval Lower	95% Confidence Interval Upper
Structure	EVA	0.018	0.894	**2.176**	**19**	**0.042**	**4.557**	**2.094**	**0.174**	**8.940**
	EVNA			2.177	18.834	0.042	4.557	2.093	0.173	8.941
Architecture	EVA	0.007	0.935	**-5.043**	**19**	**0.000**	**-11.906**	**2.361**	**-16.848**	**-6.965**
	EVNA			-5.043	18.801	0.000	-11.906	2.361	-16.852	-6.961
Technical Installations	EVA	1.459	0.242	**4.970**	**19**	**0.000**	**9.801**	**1.972**	**5.673**	**13.929**
	EVNA			5.070	17.367	0.000	9.801	1.933	5.729	13.873
Site Overheads	EVA	3.285	0.086	*-1.802*	*19*	*0.087*	*-1.725*	*0.957*	*-3.727*	*0.278*
	EVNA			-1.866	13.814	0.083	-1.725	0.924	-3.710	0.261
Unit Cost	EVA	0.941	0.346	*-2.042*	*17*	*0.057*	*-86.404*	*42.314*	*-175.679*	*2.871*
	EVNA			-2.174	16.949	0.044	-86.404	39.749	-170.286	-2.522
Initial Unit Price	EVA	0.174	0.681	**-2.222**	**18**	**0.039**	**-100.576**	**45.273**	**-195.692**	**-5.460**
	EVNA			-2.222	17.920	0.039	-100.576	45.273	-195.722	-5.430
Final Unit Price	EVA	0.054	0.821	-1.412	12	0.183	-106.575	75.453	-270.974	57.823
	EVNA			-1.443	11.650	0.175	-106.575	73.854	-268.029	54.878

EVA - Equal variances assumed
EVNA - Equal variances not assumed

The unit cost and initial price are also statistically different between residential and office buildings, if a 10% threshold is considered for the unit cost. The same is not verified for the final cost, but this can be attributed to the combination of the cost deviations and, mostly, to the smaller sample of project with final price data available. The bootstrapping results (not presented herein the full table of results) confirms the results obtained for the parameters (unit cost, initial or final price), with the unit cost difference closer to be statistically significant at a 5% significance level (p-value= 0.055).

It is interesting to notice that the total cost and prices (initial and final) of office buildings are slightly higher than for residential buildings, but the unit cost and prices are slightly lower. This implies that the office buildings in the sample are larger, in average, than the residential buildings, but that the lower expenses on architecture are only partially compensated by the more expensive structure and technical installations.

Table 4 reveals the statistical significance of the influence of the 2008 economic crisis and the international bailout that followed until 2014 on the unit cost and prices of the projects of the office building projects. Within the residential buildings in Portugal, only 2 were executed between 2008 and 2015 (in 2014 and 2015). As such, it is impossible to assess the influence of this exogenous variable on the financial performance of the residential building projects in separate. Considering all projects, the unit cost difference during the crisis is no longer statistically significant and the final cost is only significant for a 10% significance level. However, this may result from the masking effect of mixing residential and office building projects and differences in sample size for cost and initial and final price. In general, the significance level with bootstrapping decreased for all the projects analyzed together and increased for the office buildings (not presented herein the full table of results). This made the unit cost difference become statistically significant for a 10% significance level (p-value=0.096). Regarding the office buildings, this made the site overheads and the unit cost difference of office buildings lose their statistical significance.

Table 4 – Means comparison between the projects developed during the economic crisis years and during the other years

Variables		Levene's test		t-test			Difference			
		F	Sig.	t	df	Sig. (2-tailed)	Mean	Std. Error	95% Confidence Interval	
									Lower	Upper
All buildings										
Structure	EVA	2.616	0.122	-0.653	19	0.521	-1.924	2.944	-8.085	4.238
	EVNA			-0.445	3	0.683	-1.924	4.325	-14.782	10.935
Architecture	EVA	0.026	0.874	1.349	19	0.193	5.919	4.387	-3.262	15.100
	EVNA			1.187	4	0.301	5.919	4.988	-7.919	19.757
Technical Installations	EVA	3.737	0.068	-1.141	19	0.268	-4.199	3.680	-11.901	3.504
	EVNA			-2.048	17	0.056	-4.199	2.050	-8.523	0.126
Site Overheads	EVA	1.252	0.277	-0.203	19	0.841	-0.268	1.316	-3.021	2.486
	EVNA			-0.322	11	0.753	-0.268	0.832	-2.090	1.554
Unit Cost	EVA	0.055	0.818	1.566	17	0.136	83.700	53.461	-29.092	196.492
	EVNA			1.610	5	0.169	83.700	51.981	-50.578	217.979
Initial Unit Price	EVA	0.290	0.597	**2.396**	**18**	**0.028**	**133.254**	**55.626**	**16.388**	**250.119**
	EVNA			2.392	5	0.066	133.254	55.701	-13.522	280.029
Final Unit Price	EVA	0.938	0.352	*1.959*	*12*	*0.074*	*152.192*	*77.701*	*-17.105*	*321.488*
	EVNA			2.284	8	0.052	152.192	66.628	-1.443	305.826
Office Buildings										
Structure	EVA	1.605	0.241	-1.536	8	0.163	-4.719	3.072	-11.802	2.364
	EVNA			-2.222	8	0.058	-4.719	2.124	-9.633	0.195
Architecture	EVA	3.441	0.101	1.216	8	0.259	4.424	3.638	-3.966	12.813
	EVNA			1.703	8	0.127	4.424	2.597	-1.566	10.414
Technical Installations	EVA	3.395	0.103	0.829	8	0.431	2.014	2.431	-3.592	7.620
	EVNA			0.980	6	0.366	2.014	2.055	-3.049	7.078
Site Overheads	EVA	4.993	0.056	**-2.723**	**8**	**0.026**	**-1.719**	**0.631**	**-3.175**	**-0.263**
	EVNA			-2.075	2	0.148	-1.719	0.828	-4.695	1.257
Unit Cost	EVA	6.878	0.039	2.612	6	0.040	98.267	37.628	6.195	190.340
	EVNA			**3.385**	**5**	**0.022**	**98.267**	**29.034**	**21.891**	**174.643**
Initial Unit Price	EVA	10.343	0.012	3.140	8	0.014	150.429	47.907	39.956	260.901
	EVNA			**4.614**	**8**	**0.002**	**150.429**	**32.605**	**74.662**	**226.195**
Final Unit Price	EVA	2.021	0.228	*2.465*	*4*	*0.069*	*181.754*	*73.733*	*-22.961*	*386.469*
	EVNA			2.465	3	0.088	181.754	73.733	-49.712	413.220

EVA - Equal variances assumed
EVNA - Equal variances not assumed

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2021 doi:10.20944/preprints202110.0201.v1

The economic crisis impacted more severely on labor cost (there was a high unemployment and salary cuts) than on materials and equipment (a portion are imported and subject to less devaluation). This is consistent with the statistical significance of the site overheads on the office building projects, considering that a large portion of the cost in this category is due to the management team.

Since the majority of the data was found to be normally distributed based on the Shapiro-Wilk test (the non-normally distributed variables were the site overheads, margin and the underground and above ground floors), the Pearson correlation was used. The results (Table 5) reveal the expected correlation between the cost and prices with the areas and between the areas and the weight of the structure. Some less obvious results include the negative correlation between the unit cost and prices and the underground area, total area and area ratio. However, this is logic since the underground areas tend to be for parking spaces, with lower demands for architecture (and technical installations works) that justify lower unit cost and prices compared to the areas above ground. The negative relation between the unit cost and price and the total may indicate the existence of a scale effect. The bootstrap results confirm the correlations (not presented herein the full table of results). For instance, the 95% confidence interval of the correlation between the total cost and the above ground area is estimated to be between 0.705 and 0.980.

For the variables that are not normally distributed, the non-parametric Spearman correlation was also used (not presented herein), leading to similar results. The exception was a positive statistically significant correlation between the number of floors above ground and the weight of the architecture costs.

The previous unidimensional statistical analysis provides some insight on the data, but fails to account for the potential interaction between the variables. In fact, a comparison of mean assumes that all the projects in each category are identical regarding all other variables and the same applies for the correlation between two variables. Since all projects are distinct amongst them, modeling the data with multiple linear regression allows identifying the independent variables that are statistically significant to explain the dependent variable, while controlling for the influence of the other independent variables variability. This approach has its own limitations, namely the fact that a linear relation and specific relation (sum) of the variables is assumed.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2021 doi:10.20944/preprints202110.0201.v1

Table 5 –Pearson correlation results

Variables		Structure	Architecture	Technical Installations	Site Overheads	Total Cost	Initial Price	Final Price	Unit Cost	Initial Unit Price	Final Unit Price	Cost Deviation
Structure	Correlation		-,425**	0.005	-0.135	,340*	,417*	0.376	-0.270	-0.153	-0.221	-0.013
	Sig. (2-tailed)		0.007	0.976	0.397	0.042	0.010	0.062	0.107	0.347	0.273	0.951
	N		21	21	21	19	20	14	19	20	14	13
Architecture	Correlation			-,543**	0.053	-0.216	-0.253	-0.143	0.322	0.253	0.319	-0.077
	Sig. (2-tailed)			0.001	0.739	0.196	0.119	0.477	0.054	0.119	0.112	0.714
	N			21	21	19	20	14	19	20	14	13
Technical Installations	Correlation				-0.302	0.205	0.221	-0.011	-0.216	-0.200	-0.209	-0.128
	Sig. (2-tailed)				0.057	0.221	0.173	0.956	0.196	0.218	0.298	0.542
	N				21	19	20	14	19	20	14	13
Site Overheads	Correlation					-,413*	-,483**	-0.331	0.012	0.005	-0.044	0.297
	Sig. (2-tailed)					0.014	0.003	0.100	0.944	0.974	0.826	0.160
	N					19	20	14	19	20	14	13
Underground Floors	Correlation	0.336	0.062	-0.109	-0.314	0.108	0.280	0.238	-0.088	-0.140	-0.089	0.149
	Sig. (2-tailed)	0.079	0.744	0.568	0.102	0.596	0.142	0.296	0.664	0.463	0.695	0.514
	N	18	18	18	18	16	18	13	16	18	13	13
Above Ground Floors	Correlation	-0.286	,380*	-0.063	-0.089	-0.090	-0.064	0.082	0.008	-0.049	0.151	0.162
	Sig. (2-tailed)	0.106	0.031	0.719	0.615	0.638	0.726	0.696	0.966	0.785	0.469	0.456
	N	19	19	19	19	17	18	14	17	18	14	13
Total Floors	Correlation	-0.098	0.327	-0.132	-0.183	-0.036	0.021	0.211	-0.072	-0.104	0.053	0.184
	Sig. (2-tailed)	0.587	0.069	0.461	0.313	0.853	0.907	0.325	0.710	0.561	0.806	0.389
	N	18	18	18	18	16	18	13	16	18	13	13
Underground Area	Correlation	,558**	-,495**	0.286	-0.273	,579**	,663**	,473*	-,520**	-,389*	-,516*	-0.051
	Sig. (2-tailed)	0.000	0.002	0.070	0.085	0.001	0.000	0.019	0.002	0.016	0.010	0.807
	N	21	21	21	21	19	20	14	19	20	14	13
Above Ground Area	Correlation	,431**	-0.100	-0.005	-,327*	,739**	,691**	,758**	-0.246	-0.216	-0.231	0.000
	Sig. (2-tailed)	0.007	0.526	0.976	0.040	0.000	0.000	0.000	0.141	0.183	0.250	1.000
	N	21	21	21	21	19	20	14	19	20	14	13
Total Area	Correlation	,539**	-,362*	0.171	-,350*	,743**	,800**	,714**	-,427*	-,358*	-,407*	0.000
	Sig. (2-tailed)	0.001	0.022	0.277	0.027	0.000	0.000	0.000	0.011	0.027	0.043	1.000
	N	21	21	21	21	19	20	14	19	20	14	13
Area Ratio	Correlation	,539**	-,362*	0.171	-,350*	,743**	,800**	,714**	-,427*	-,358*	-,407*	0.000
	Sig. (2-tailed)	0.001	0.022	0.277	0.027	0.000	0.000	0.000	0.011	0.027	0.043	1.000
	N	21	21	21	21	19	20	14	19	20	14	13

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2021 doi:10.20944/preprints202110.0201.v1

The cost and prices, both total and unit, were selected as independent variables, along with the cost deviation. All other variables were considered as potential predictors. A hybrid approach was used to select the predictors to include in the models, combining expert judgment and the best subsets tool with the Akaike Information Criterion. The option for this hybrid approach resulted from an experimental stage using only statistical tools to select the predictors (stepwise and best subsets using Akaike Information Criterion, Ajusted R2 and Overfit Prevention Criterion) produced models with very high fit, but not robust from an engineering point of view. Furthermore, the models for predicting total cost and price were developed without intercept to ensure that the value tends to zero when the project size decreases. There were no signs of heteroscedasticity (White and Breusch-Pagan tests), non-normal distribution of the residuals (Shapiro-Wilk test) or influential observations (Cook's distance) in all hybrid models. Still, robust standard errors were used in all models. There is also no evidence of specification problems (linktest), and the functional forms seem appropriate (Ramsey test).

The regression models for the initial and final price model are presented in Table 6. The R2 of the models is 0.92. Given the high R2 obtained, the models with the predictors selected with statistical tools alone produced similar results in terms of fit to the data. For instance, using the best subsets with the adjusted R2 as criterion to select variables it was possible to obtain a model for the initial price with an R2 of 0.95 using the following variables: i) area above ground; ii) area x type; iii) floors above ground; iv) total floors; and v) area ratio. However, this comes with a cost in terms of outliers (3 cases were identified as outliers using the Cook's distance) and represents a potential overfit (a model with 5 variables for a dataset with 18 cases). Due to the reduced size of the sample available (8 residential and 6 office buildings) for developing the final price model, the result should be looked with due care.

Table 6 – Regression models for the initial and final price

Parameter	B	Robust Std. Error[a]	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Initial Price						
Above Ground Area	735.860	138.565	5.311	0.000	443.512	1028.207
Underground Area	462.428	121.467	3.807	0.001	206.155	718.701
Area X Crisis	-102.426	36.276	-2.824	0.012	-178.961	-25.890
Final Price						
Above Ground Area	1393.707	399.891	3.485	0.005	513.554	2273.860
Underground Area	232.331	127.608	1.821	0.096	-48.531	513.194
Area X Type	-181.507	118.842	-1.527	0.155	-443.077	80.062

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2021 doi:10.20944/preprints202110.0201.v1

Due to the confidentiality, the model for the total cost cannot be disclosed. The variables in the models were the same of the initial price models, which is logic since the difference between both is the margin set by the contractor. However, the results of the model are depicted in Figure 2, corresponding to an R2 of 0.94.

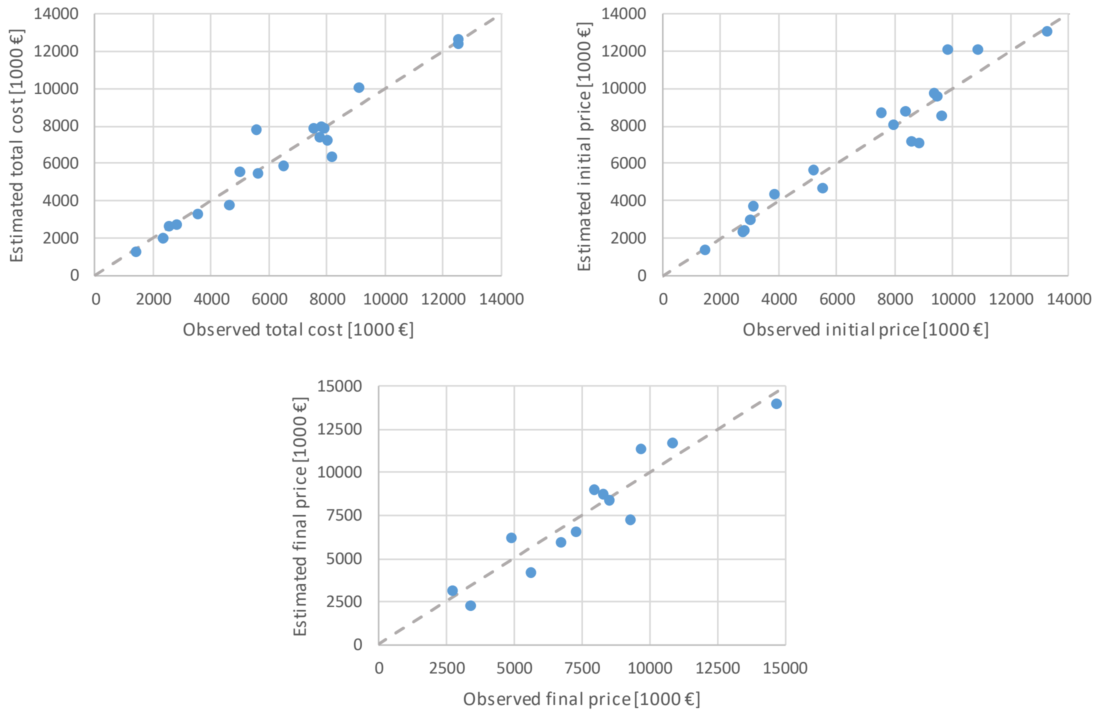

Figure 2 – Observed versus estimated total cost and initial and final price

Both total and unit cost or prices are connected, but the high correlation between the total cost or price and the construction area may mask the influence of other variables. Considering the confidentiality issues and the limitations of sample size, only the initial unit price was modelled. The first model obtained attained an R2 of 0.505 using as predictors the variables: i) floors above ground; ii) total floors; iii) floor ratio; and iv) economic crisis.

However, since a clear non-linear pattern was visible when plotting observed versus predicted initial unit prices, a non-linear multiple regression model was developed. The non-linearity was accounted for by including power coefficients in the scale predictors. The best model resulted in a power of 1.011 for the floors above ground and 1.608 for the total floors, increasing the R2 to 0.720 (Table 7).

There is influence of the economic crisis, but the proportion of underground and above ground floors became statistically significant with the removal of the area from the model. The difference between the linear and non-linear models can be observed in Figure 3, evidencing the fit increase in the later.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2021 doi:10.20944/preprints202110.0201.v1

Table 7 – Regression models for the initial unit price

Parameter	B	Robust Std. Error[a]	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	503.309	36.238	13.889	0.000	425.022	581.596
Above Ground Floors $^{1.011}$	-160.284	30.403	-5.272	0.000	-225.966	-94.602
Total Floors $^{1.608}$	17.286	3.129	5.524	0.000	10.525	24.046
Floor Ratio	117.935	25.915	4.551	0.001	61.949	173.920
Economic Crisis=0	211.752	36.914	5.736	0.000	132.005	291.499
Economic Crisis=1	0.000					

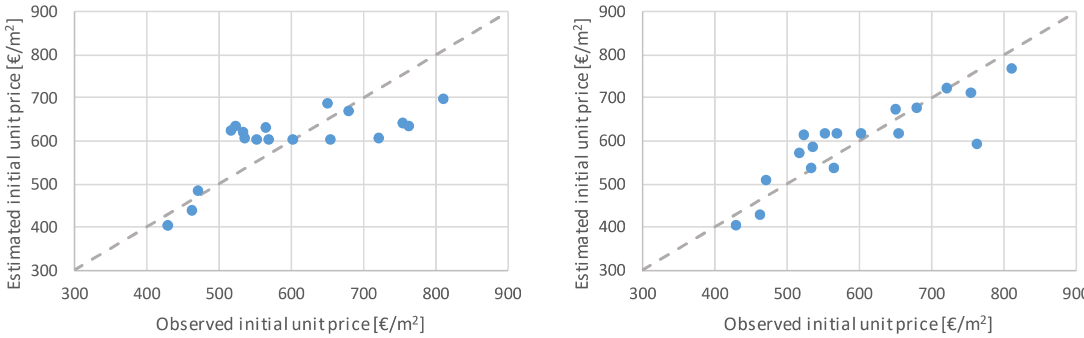

Figure 3 – Observed versus estimated initial unit price (left: linear model; right: non-linear model)

The apparently lower fit of the models for the unit price is misleading. In fact, multiplying the area by the initial unit prices estimated with the non-linear model to determine the total initial price achieves an R2 of 0.97 (Figure 4). This fit difference between the models for the total and unit prices results from the correlation between the total area and the number of floors. This correlation produces multicollinearity between the variables, resulting in the exclusion of the number of floors from any model in which the area is also used. Removing the influence of the area by modelling the unit price allows for the influence of the number of floors to be accounted for, which explains the accuracy increase.

Bootstrapping was also used in the development of the regression models and confirm the statistical significance of the regression coefficients for a 95% confidence interval. Generally, the significance of the regression coefficients decreased, but the p-value remained lower than 0.05 in all cases except for the final price model. For this model, the regression coefficients of the Underground Area and Area X Type already exceeded the 5% significance threshold even without bootstrapping, which can be attributed to the small number of projects for which the final price was available.

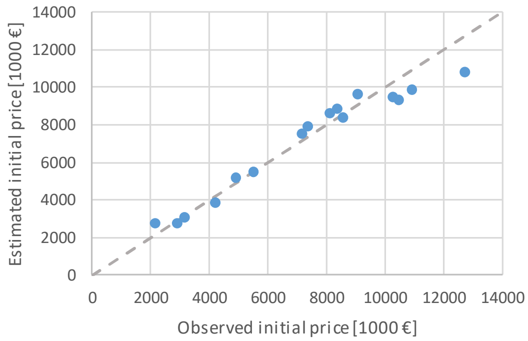

Figure 3 – Observed versus estimated initial price using the non-linear initial unit price model

With the purpose of testing and validating the models developed in this research, the model for the initial price was applied to a project currently under development by the organization. Considering that the project used for validation was estimated in over 45 million euros, significantly higher than the projects in the dataset, and that the difference to the price estimated by the organization was less than 5%, there was a positive feedback from the organization regarding the accuracy and the extrapolation capability of the model.

In the sample of 13 projects (6 office and 7 residential) for which initial and final prices were available, an average cost deviation of 3.5% was obtained. Only 3 projects had a final price lower than the initial estimate (average of -6.5%). The projects with positive cost deviations were, in average, 6.5% costlier and there was no project without cost deviation. Comparing with the literature available, which generally adopts the owner perspective, the magnitude of the cost deviation is clearly smaller than usually reported and it becomes evident that the contractor always experiences some cost deviation, even if that is not reflected on the bill of the owner.

Either due to the limitations of the dataset, the fact that the projects are limited in type, the spatial context and stakeholders involved, or a combination of these and other factors, the cost deviation depend on specific aspects of each project that are not captured by the general information used herein and it was not possible to model them. The only statistically significant result obtained was the high Person correlation (0.814) between the number of underground floors and the cost deviation of office buildings. The corresponding regression model indicates that the average cost deviation in office buildings increase 0.65% per underground floor, but this was obtained from a sample of only 6 projects and its validity is questionable.

5 CONCLUSIONS

This research revisits the topic of cost estimation and deviation of construction projects, but adopting an innovative perspective of a contractor, which seems uncommon from the literature review carried out.

Furthermore, to the best of our knowledge, this is one of the few efforts linking endogenous and exogenous variables in cost estimation functions.

Contrarily to most research available, only similar projects (premium residential and office buildings) from a single promotor-contractor are used. This compromises the size of the database available, but eliminates the variability of cost estimates and deviations due to: i) factors related to the contractor or the designer (e.g., experience; competencies; dimension; management models); ii) characteristics of the projects (e.g., premium buildings, social buildings, public buildings); iii) relation between owner and contractor (e.g., type of owner – public, private; type of contract - design-bid-build, design-build; payment method – lump sum, unit prices); and iii) aspects associated with the location (e.g., weather conditions; laws and regulations). Since the projects are promoted by the real estate company of the same group, the commercial strategy issues related to the degree of competition of the market has less effect on the cost of the projects. The contract does not have to adjust its margin to win the contract and so the influence of the level of competition in the market is only limited to the portion of the project that is executed by subcontractors. By doing so, the results presented herein grasp the "real" cost estimation and deviations driven by project related factors. The high accuracy of the cost estimation m The results obtained with these restrictions support the importance of the technical expertise of the involved parties in the cost estimation and deviations reported in the literature. Comparing the average and range of the cost deviations in this study with other authors, it is licit to assume that, at least, a portion of the difference is due to the experience of the teams involved and not only due to project (e.g., construction technology) or context (e.g., weather conditions) specificities. Other factor possibly underlying the differences in terms of the magnitude of the cost deviations is the collaborative effort of promoter and contractor in this case, reducing the conflicts that are not rare in the traditional design-bid-build contracts where the promoter has limited expertise/resources regarding the execution stage of the construction project.

Despite the reduced sample size when compared to other studies, it is noticeable that the cost deviations in this context are smaller than what is typically reported when adopting the owner, either public or private, perspective. The generalization of the results may be limited, but they do provide a source for other contractors benchmark their performance and the methodology proposed sets a basis for developing similar studies both in research or practical contexts. In fact, the linear and non-linear regression models developed are of easy interpretation and assessment from an expert, which was done with good results, whereas artificial intelligence models are black-boxes impossible or very difficult to be validated by experts. The practical expert validation carried out, along with the bootstrapping results, reinforce the applicability of the models for the specific context in which was developed and corroborates the applicability of the methodology in other contexts.

The models developed for estimating cost have a very high fit to the data and highlight the influence of the economic crisis and international bailout on the construction costs. In Portugal, the price of construction projects in open competition also suffered a strong reduction during this period due to the lack of both private and public construction projects. However, since the price is driven not only by the cost but also by the market conditions (e.g., relation between demand and supply), the variation is not necessarily identical, and this research is able to capture the pattern of the cost.

The cost deviations seem to depend more in particular aspects of each project than overall characteristics, despite the positive statistically significant relation between the number of underground floors and the cost deviations in office buildings found.

Data Availability Statement

Some or all data, models, or code generated or used during the study are proprietary or confidential in nature and may only be provided with restrictions. The estimated cost and initial and final values of each individual project can only be disclosed normalized format (actual value divided by the sample average). The model for the estimated cost cannot be disclosed.

REFERENCES

Abusafiya, H.A.M., and S.M.A. Suliman. 2017. "Causes and Effects of Cost Overrun on Construction Project in Bahrain: Part I (Ranking of Cost Overrun Factors and Risk Mapping)." *Mod. Appl. Sci.*, 11(7), 20. https://doi.org/10.5539/mas.v11n7p20

Ahn, J., Ji, S.-H. Ji, S.J. Ahn, M. Park, H.-S. Lee, N. Kwon, E.-B. Lee and Y. Kim. 2020. "Performance evaluation of normalization-based CBR models for improving construction cost estimation." *Autom. Constr.*, 119: 103329. https://doi.org/10.1016/j.autcon.2020.103329

Al-Tabtabai, H., A.P. Alex and M. Tantash. 1999. "Preliminary cost estimation of highway construction using neural networks." *Cost Eng.*, 41(3), 19–24.

Annamalaisami, C.D. and Kuppuswamy, A. 2019. "Reckoning construction cost overruns in building projects through methodological consequences." International Journal of Construction Management, https://doi.org/10.1080/15623599.2019.1683689

Attalla, M. and T. Hegazy. 2003. "Prediction cost deviation in reconstruction project: Artificial neural networks versus regression." *J. Constr. Eng. Manage.*, 129:4, 405–411. https://doi.org/10.1061/(ASCE)0733-9364(2003)129:4(405)

Balali, A., Moehler, R.C. and Valipour, A. 2020. "Ranking cost overrun factors in the mega hospital construction projects using Delphi-SWARA method: an Iranian case study." International Journal of Construction Management. https://doi.org/10.1080/15623599.2020.1811465

Bayram, S., M.E. Ocal, E.L. Oral and C.D. Atis. 2016. "Comparison of multi layer perceptron (MLP) and radial basis function (RBF) for construction cost estimation: the case of Turkey." *J. Constr. Eng. Manage.*, 22(4), 480-490. https://doi.org/10.3846/13923730.2014.897988

Bucciol, A., Chillemi, O. and Palazzi, G. 2013. Cost overrun and auction format in small size public works. European Journal of Political Economy, 30, 35–42. https://doi.org/10.1016/j.ejpoleco.2013.01.002

Catalão, F.P., C.O. Cruz, and J.M. Sarmento. 2019a. "Exogenous determinants of cost deviations and overruns in local infrastructure projects." *Constr. Manag. Econ.*, 37(12), 697–711. https://doi.org/10.1080/01446193.2019.1576915

Catalão, F.P., C.O. Cruz, and J.M. Sarmento. 2019b. "The determinants of cost deviations and overruns in transport projects, an endogenous models approach." *Transp. Policy*, 74, 224-238. https://doi.org/10.1016/j.tranpol.2018.12.008.

Catalão, F.P., C.O. Cruz, and J.M. Sarmento. 2020. Public management and cost overruns in public projects. *Int. Public Manag. J.*, https://doi.org/10.1080/10967494.2020.1804498

Cruz, C.O., and Branco, F. (2020). Reconstruction Cost Model for Housing Insurance. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 12(3), 05020007. https://doi.org/10.1061/(ASCE)LA.1943-4170.0000408

Derakhshanalavijeh, R. and J.M.C. Teixeira. 2017. "Cost overrun in construction projects in developing countries, Gas-Oil industry of Iran as a case study." *J. Civ. Eng. Manag.*, 23(1), 125–136. https://doi.org/10.3846/13923730.2014.992467

Doğan, S.Z., D. Arditi and H.M. Günaydin. 2006. "Determining attribute weights in a CBR model for early cost prediction of structural system." *J. Constr. Eng. Manag.*, 132(10), 1092-1098. http://dx.doi.org/10.1061/(ASCE)0733-9364(2006)132:10(1092)

Elhag, T.M.S., A.H. Boussabaine and T.M.A. Ballal. 2005. “Critical Determinants of construction tendering costs: Quantity Surveyor's standpoint.” *Int. J. Proj. Manag.*, 23, 538-545. https://doi.org/10.1016/j.ijproman.2005.04.002

Elhag, T.M.S., A.H. Boussabaine. 1998. “An artificial neural system for cost estimation of construction projects.” In *Proc. 14th Annual ARCOM Conf*., 219–226, Reading, U.K..

Emsley, M.E., D.J. Lowe, A.R. Duff, A. Harding and A. Hickson. 2002. “Data modeling and the application of a neural network approach to the prediction of total construction costs.” *Constr. Manage. Econom*., 20(6), 465–472. https://doi.org/10.1080/01446190210151050

Flyvbjerg, B., Ansar, A., Budzier, A., Buhl, S., Cantarelli, C., Garbuio, M., Glenting, C., Holm, M.S., Lovallo, D., Molin, E., Rønnest, A., Stewart, A., van Wee, B. 2019. On de-bunking “Fake News” in the post-truth era: How to reduce statistical error in research. Transp. Res. Part A Policy Pract., 126, 409–411. https://doi.org/10.1016/j.tra.2019.06.011

Flyvbjerg, B., Ansar, A., Budzier, A., Buhl, S., Cantarelli, C., Garbuio, M., Glenting, C., Holm, M.S., Lovallo, D., Molin, E., Rønnest, A., Stewart, A., van Wee, B. 2018. Five things you should know about cost overrun. Transport. Res. Part A: Policy Pract., 118, 174-190. https://doi.org/10.1016/j.tra.2018.07.013

Flyvbjerg, B., C. Hon and W.H. Fok. 2016. “Reference class forecasting for Hong Kong’s major roadworks projects.” *Proc. Inst. Civ. Eng. Civ. Eng*., 169(6), 17–24. https://doi.org/10.1680/jcien.15.00075.

Flyvbjerg, B., Holm, M. K. S. and Buhl, S. L. 2004. What causes cost overrun in transport infrastructure projects? Transport Reviews, 24(1), 3-18. https://doi.org/10.1080/0144164032000080494a

Flyvbjerg, B., Holm, M. S. and Buhl, S. 2002. Underestimating costs in public works projects. Error or lie? Journal of the American Planning Association, 68(3), 279-295. https://doi.org/10.1080/01944360208976273

Günaydin, H.M. and S.Z. Doğan. 2004. “A neural network approach for early cost estimation of structural systems of buildings.” *Int. J. Project Manage*., 22(7), 595–602. https://doi.org/10.1016/j.ijproman.2004.04.002.

Gunduz, M., L.O. Ugur and E. Ozturk. 2011. “Parametric cost estimation system for light rail transit and metro trackworks.” *Expert Syst. Appl*., 38(3), 2873–2877. https://doi.org/10.1016/j.eswa.2010.08.080.

Hegazy, T. and A. Ayed. 1998. “Neural network model for parametric cost estimation of highway projects.” *J. Constr. Eng. Manage.*, 124:3, 210–218. https://doi.org/10.1061/(ASCE)0733-9364(1998)124:3(210)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2021 doi:10.20944/preprints202110.0201.v1

Hu, X., B. Xia, M. Skitmore and Q. Chen. 2016. "The application of case-based reasoning in construction management research: An overview." *Autom. Constr.*, 72(2), 65-74. https://doi.org/10.1016/j.autcon.2016.08.023.

Ji, C., T. Hong and C. Hyun. 2010. "CBR revision model for improving the cost prediction accuracy in the multifamily housing project." *J. Manage. Eng.*, 26(4), 229–236. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000018

Jin, R., K. Cho, C. Hyun and M. Son. 2012. "MRA-based revised CBR model for cost prediction in the early stage of construction projects." *Expert Syst. Appl.*, 39(5), 5214–5222. https://doi.org/10.1016/j.eswa.2011.11.018.

Jin, R., S. Han, C. Hyun and J. Kim. 2014. "Improving accuracy of early stage cost estimation by revising categorical variables in a case-based reasoning model." *J. Constr. Eng. Manage.*, 140(7): 04014025. https://doi.org/10.1061/(asce)co.1943-7862.0000863

Jin, R., S. Han, C. Hyun and Y. Cha. 2016. "Application of Case-Based Reasoning for Estimating Preliminary Duration of Building Projects." *J. Constr. Eng. Manage.*, 142(2): 04015082. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001072

Kaming, P.F., Olomolaiye, P.O., Holt G.H. and Harris, F.C. 1997. "Factors influencing construction time and cost overruns on high-rise projects in Indonesia." Construction Management & Economics, 15(1), 83-94. https://doi.org/10.1080/014461997373132

Karaca, I., D.D. Gransberg and H.D. Jeong. 2020. "Improving the Accuracy of Early Cost Estimates on Transportation Infrastructure Projects." *J. Manage. Eng.*, 36(5): 04020063. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000819.

Kim, G., D. Seo and K. Kang. 2005. "Hybrid models of neural networks and genetic algorithms for predicting preliminary cost estimates." *J. Comput. Civ. Eng.*, 19(2), 208–211. https://doi.org/10.1061/(ASCE)0887-3801(2005)19:2(208)

Kim, G., S. An and K. Kang. 2004. "Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning." *Build. Environ.*, 39(10), 1235–1242. https://doi.org/10.1016/j.buildenv.2004.02.013

Kim, K.J. and K. Kim. 2010. "Preliminary cost estimation model using case-based reasoning and genetic algorithms." *J. Comput. Civ. Eng.*, 24(6), 499-505. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000054

Li, H., Q.P. Shen and P.E.D. Love. 2005. “Cost modelling of office buildings in Hong Kong: an exploratory study.” *Facilities*, 23(9), 438-452. https://doi.org/10.1108/02632770510602379

Love, P.E.D. and Ahiaga-Dagbui D.D. 2018. De-bunking ‘fake news’ in a post-truth era: the plausible untruths of cost underestimation in transport infrastructure projects Transport. Res. A: Policy Pract., 113, 357-368. https://doi.org/10.1016/j.tra.2018.04.019

Love, P.E.D., Ika, L. and Ahiaga-Dagbui D.D. 2019 On de-bunking ‘fake news’ in a post-truth era: why does the Planning Fallacy explanation for cost overruns fall short? Transport. Res. A: Policy Practice., 126, 397-408. https://doi.org/10.1016/j.tra.2019.06.013

Love, P.E.D., Sing, C.-P., Carey, B., and Kim, J.-T. (2015). Estimating construction contingency: accommodating the potential for cost overruns in road construction projects. Journal of Infrastructure Systems, 21(2), 04014035. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000221

Love, P.E.D., Sing, C.-P., Wang, X., and Tiong, R. (2013). Determining the probability of cost overruns in Australian construction and engineering projects. Journal of Construction Engineering and Management, 139(3), 321–330. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000575

Love, P.E.D., Sing, X. C.P., & Tiong, R. L. K. 2020. Determining the Probability of Project Cost Overruns. Journal of Construction Engineering and Management, 146(6), 04020060. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001851

Lowe, D.J., M.W. Emsley and A. Harding. 2006. “Predicting construction cost using multiple regression techniques.” *J. Constr. Eng. Manage*., 132(7), 750–758. https://doi.org/10.1061/(ASCE)0733-9364(2006)132:7(750)

Muratova, A. and I. Ptukhina. 2020. “BIM as an Instrument of a Conceptual Project Cost Estimation.” In *Proc. of EECE 2019, Lecture Notes in Civil Engineering*, vol 70. Springer, Cham. https://doi.org/10.1007/978-3-030-42351-3_41

Pearl, R. G., Akintoye, A., Bowen, P. A. and Hardcastle, C. 2003. Analysis of tender sum forecasting by quantity surveyors and contractors in South Africa. Journal for the physical and development sciences: Acta Stuctilla, 10(1), 5-35.

Picken, D.H. and B.D. Ilozor. 2003. “Height and construction costs of buildings in Hong Kong.” *Constr. Manage. Econom*., 21(2), 107–111. https://doi.org/10.1080/0144619032000079671

Rafiei, M. and H. Adeli. 2018. “Novel machine-learning model for estimating construction costs considering economic variables and indexes.” *J. Constr. Eng. Manage.*, 144(12): 04018106. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001570.

Reyes, J. and Mansfield, J. 2001. The assessment of risk in conservation refurbishment projects. Structural Survey, 19(5), 238–244. https://doi.org/10.1108/02630800110412480

Shehu, Z., I.R. Endut, A. Akintoye and G.D. Holt. 2014. “Cost overrun in the Malaysian construction industry projects: A deeper insight.” *Int. J. Proj. Manag.*, 32(8), 1471–1480. https://doi.org/10.1016/j.ijproman.2014.04.004

Shrestha, P.P. and Fathi, M. 2019. “Impacts of Change Orders on Cost and Schedule Performance and the Correlation with Project Size of DB Building Projects. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 11(3), 04519010. https://doi.org/10.1061/(ASCE)LA.1943-4170.0000311

Skitmore, R.M. and S.T. Ng. 2003. “Forecast models for actual construction time and cost.” *Build. Environ.*, 38(8), 1075–1083. https://doi.org/10.1016/S0360-1323(03)00067-2

Sonmez, R. 2004. “Conceptual cost estimation of building projects with regression analysis and neural networks.” *Can. J. Civ. Eng.*, 31(4), 677–683. https://doi.org/10.1139/l04-029

Soto, B.G., T. Streule, M. Klippel, O. Bartlomé and B.T. Adey. 2020. “Improving the planning and design phases of construction projects by using a Case-Based Digital Building System.” *Int. J. Constr. Manag.*, 20:8, 900-911. https://doi.org/10.1080/15623599.2018.1502929

Sousa, V. and I. Meireles. 2018. “The influence of the construction technology in time-cost relationships for sewerage projects in the U.S..” *Water Resour. Manag.*, 32(8), 2753-2766. https://doi.org/10.1007/s11269-018-1956-y

Sousa, V., N.M. Almeida and L. Alves Dias. 2014b. “Role of statistics and engineering judgment in developing optimized time-cost relationship models.” *J. Constr. Eng. Manage.*, 140(8): 04014034. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000874

Sousa, V., N.M. Almeida, L. Alves Dias and F. Branco. 2014a. “Risk-informed time-cost relationship models for sanitation projects.” *J. Constr. Eng. Manage.*, 140(5): 06014002. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000848

Stoy, C. and H.-R. Schalcher. 2007. “Residential Building Projects: Building Cost Indicators and Drivers.” *J. Constr. Eng. Manage.*, 133(2), 139-145. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000874

Stoy, C., S. Pollalis and H.-R. Schalcher. 2008. "Drivers for Cost Estimating in Early Design: Case Study of Residential Construction." *J. Constr. Eng. Manage*., 134(1), 32-39. https://doi.org/10.1061/(ASCE)0733-9364(2008)134:1(32)

Swei, O., J. Gregory and R. Kirchain. 2017. "Construction cost estimation: A parametric approach for better estimates of expected cost and variation." *Transport. Res. B-Meth*., 101, 295-305. https://doi.org/ 10.1061/(ASCE)0733-9364(2007)133:2(139)

Sweis, G.J., R. Sweis, M.A. Rumman, R.A. Hussein and S.E. Dahiya. 2013. "Cost overruns in public construction projects: the case of Jordan." *J. Am. Sci*., 9(7), 134-141. https://doi.org/110.5539/ijbm.v8n23p120

Thalmann, P. 1998. "A low-cost construction price index based on building functions." In *Proc. 15th Int. Cost Engineering Congress*, ICEC, Rotterdam, The Netherlands.

Trost, S.M. and G.D. Oberlender. 2003. "Prediction accuracy of early cost estimates using factor analysis and multivariate regression." *J. Constr. Eng. Manage*., 129:2, 198–204. https://doi.org/10.1061/(ASCE)0733-9364(2003)129:2(198)

Wheaton, W.C. and W.E. Simonton. 2007. "The secular and cyclic behavior of 'true' construction costs." *J. Real Estate Res.,* 29(1), 1-26. https://www.jstor.org/stable/24888194