Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 278 (2026) 2054–2061

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2025

A Hybrid Simplified Approach for Cost Estimation in Construction Project Management

Natalia Trapani[a]*, Giovanni Nicoletti[a], Giorgio Platania[b]

[a]*University of Catania, via Santa Sofia, 68, 95123 Catania, Italy*
[b]*Project Manager Professional, PMP©, via Monte Lauro, 4, 95030 Tremestieri Etneo (CT), Italy*

Abstract

Accurate cost estimation is a critical success factor in construction project management, particularly in sectors where project approval and resource allocation depend on reliable forecasts. This paper presents a hybrid parametric model, designed to support project managers in the preliminary stages of project planning by providing reliable cost forecasts based on a combination of historical data, project-specific variables, and contextual factors. Validation was performed through real-world case studies from bank branch projects. The model's sector-specific calibration, based on eight years of historical data, ensures that it accurately reflects real-world cost drivers such as location, intervention type, and material cost trends. The validation demonstrates the model's effectiveness for standard-sized projects and highlights areas for further refinement. Validation was performed through five real-world case studies from bank branch projects. The approach offers a practical, scalable tool to support project managers in budgeting and decision-making, with potential for adaptation to other contexts and for digital integration in digital project management platforms. The model achieved an average prediction error band of ±7%, demonstrating its reliability for standard-sized projects.

* Corresponding author. Tel.: +39-095-738-2465
E-mail address: natalia.trapani@unict.it

1877-0509 © 2025 The Authors. Published by Elsevier B.V.

Keywords: Cost Estimation, Parametric Model, Construction Management, Digital Tools.

1. Introduction

Accurate and robust cost forecasting is essential for effective project management in construction, influencing decisions on project feasibility, resource allocation, and risk governance from the earliest stages. The increasing complexity of modern construction projects and the need for timely and informed decisions have driven the development of advanced estimation frameworks capable of delivering robust predictions even when comprehensive project information is lacking. Among cost estimation approaches, parametric models are particularly valuable due to their inherent ability to provide rapid and scalable cost projections based on a limited set of measurable project attributes. They are especially useful in early-stage planning, where time constraints and information gaps make detailed (bottom-up) or comparative (analogous) estimation impractical. However, their effectiveness depends on rigorous calibration with real-world data and demonstrated adaptability across diverse project types.

Despite advances in data-driven techniques and digital integration, there remains a notable lack of validated, sector-adaptable parametric models that combine transparency, scalability, and practical usability for project managers. This paper addresses this gap by proposing and validating a hybrid parametric cost estimation model, developed using historical data from bank branch construction and renovation projects. Bank branches projects provide an ideal case study due to their standardized requirements, comprehensive historical cost data, and emphasis on precise budgeting and regulatory compliance. By systematically analyzing key cost drivers, such as location, building age, intervention type, and operational requirements, this research develops a calibrated model designed for both accuracy and transferability.

The primary objective is to provide a practical tool for robust early-stage cost forecasts, enhancing decision-making and resource planning. While validated within the banking sector, the model's methodology is designed for scalability and adaptation to other sectors with similar characteristics, bridging the gap between model-driven and data-driven estimation paradigms and enabling future integration with digital project management platforms.

Nomenclature

BIM	Building Information Modeling	CER	Cost Estimation Relationships
EVM	Earned Value Management	FDI	Foreign Direct Investment
ISTAT	Italian National Institute for Statistical Data (*Istituto Nazionale di Statistica*)		
OMI	Real Estate Market Observatory (*Osservatorio del Mercato Immobiliare*)		

2. Literature review

The evolution of cost estimation in construction projects has evolved from traditional manual methodologies to sophisticated data-driven and hybrid approaches [1]. Historically, model-driven techniques, including bottom-up, top-down and rule-based models, relied on detailed quantity take-offs, expert judgment, and historical analogies [2]. These methods established structured relationships between project parameters and costs, offering interpretability and adaptability, particularly when historical data were scarce or project contexts vary. However, such approaches are time-consuming and often lack accuracy in early project phases due to limited detailed information.

Parametric estimation methodologies emerged to address these limitations by employing statistical relationships between project parameters (e.g., geographical area, floor surface, intervention type) and comprehensive historical cost databases. This approach enables rapid, scalable cost estimates, which are essential for early-stage planning and feasibility studies. Regression analysis and the systematic development of cost estimation relationships (CERs) are central to parametric methods, effectively bridging expert-driven and data-driven paradigms [3].

Recent advancements include the adoption of digital tools and analytics in construction cost estimation [4, 5]. The integration of Building Information Modeling (BIM), cloud-based platforms, and digital twin technologies has enabled real-time data sharing, dynamic cost management, and enhanced collaboration among stakeholders [6]. These technologies facilitate continuous updating of cost estimates as design evolves, enhancing both accuracy and responsiveness.

Data-driven approaches, including machine learning (ML) algorithms and advanced statistical analysis, leverage extensive project datasets to identify complex patterns and generate predictive models. ML has become transformative in project cost forecasting. Recent studies [7; 8] demonstrate that ML models, specifically XGBoost, outperform traditional Earned Value Management (EVM) and other ML algorithms (Random Forest, Support Vector Regression, LightGBM, CatBoost) in both prediction accuracy and forecasting timeliness. Using real data from 110 projects, XGBoost model provided reliable forecasts across all project development stages, enabling early warning signals for potential cost overruns and supporting proactive project control mechanisms. Additional research [9] highlights ML's capability to capture non-linear relationships and complex interactions frequently overlooked by traditional modeling approaches, in both conceptual cost estimation [10] as well as risk analysis [11].

The literature reveals complementary strengths and limitations of both model-driven and data-driven paradigms. ML models exhibit superior predictive accuracy when applied to large, well-structured, and high-quality datasets [12]. Moreover, their "black box" nature can hinder understanding of underlying cost drivers and may exhibit poor generalization to projects outside their training data domains. Conversely, traditional model-driven approaches offer enhanced transparency and methodological adaptability but remain relatively rigid, failing to capture emerging industry trends or context-specific characteristics not explicitly defined within their operational parameters [13].

Recent research has increasingly focused on hybrid approaches that strategically combine the strengths of model-driven and data-driven forecasting. Advanced hybrid models integrating artificial neural networks (ANNs) with decision trees (DTs) or combining Design-Build-Operate (DBO) frameworks with backpropagation neural networks (BPNN) demonstrate superior accuracy and computational efficiency compared to traditional methods [14; 15; 9]. These approaches are further enhanced by ensemble methods and the integration of uncertainty quantification, which are crucial for robust decision-making in construction projects [16]. Case studies demonstrate practical hybrid model applications in different contexts, such as pile foundation cost estimation and conceptual project cost forecasting, confirming their enhanced effectiveness in real-world scenarios [14; 15]. Recent research emphasizes the need for sector-specific cost estimation models that account for unique operational, regulatory, and contextual factors. Economic analyses reveal macroeconomic impact, such as public and private investment, foreign direct investment (FDI), and regional disparities, on construction cost structures. Pham et al. [17] empirically demonstrates that public investment and FDI significantly influence regional construction costs and economic growth patterns, while private investment encounters legal and financial constrained. These findings highlight the importance of incorporating economic context into cost estimation models, which must accommodate sector-specific requirements and regional economic conditions [18; 19].

This study addresses the following research question: 1) Can a hybrid parametric model, calibrated with sector-specific historical data and contextual factors, provide reliable early-stage cost forecasts for bank branch construction and renovation projects? 2) How robust is the model's predictive performance given a limited validation sample, and how interpretable are the model's multipliers for practical use? Accordingly, the primary challenge of this work lies in explicitly addressing the imperative for a compromise, specifically, developing a hybrid calibrated approach that strategically leverages the strengths of both data-driven and model-driven forecasting paradigms. The proposed model is distinguished by its foundation in historical data. It is constructed upon a rigorous analysis of past project data, thereby ensuring that it accurately reflects real-world cost drivers and prevailing trends. Furthermore, the model integrates explicit mechanisms to adjust for historical trends in material costs and other pertinent variables, which is crucial for maintaining the relevance and accuracy of forecasts as dynamic market conditions continue to evolve.

3. The cost estimation model

3.1. Base Cost Estimation

The main objective of this work is to create a model that allows professionals to obtain a cost estimate as accurately as possible before the project is realized, considering the history of similar projects and the types of variants that may lead to additional or reduced costs. The analysis of the documentation provided by an Italian company, operating in the construction of bank branches on the Italian territory, has made it possible to identify the main factors that impact cost estimation and that can be used to calibrate a parametric model. The starting point is the average parametric cost (C_0) per square meter for bank branches, which includes real estate investment, furniture and furnishings, and security systems. This data was obtained by using costs from completed bank branch projects in a period of eight years.

The floor surface measured in square meter indicates whether the real estate units are small, medium, or large compared to the real estate market in question. The average parametric cost of realization for banks, evaluated over eight years, is a key factor. The cost differences (measured in €/m^2) are limited to ±2.5% in the evaluation period.

3.2. The Cost Estimation Framework

The definition of the model framework involves identifying key cost drivers, collecting, and analyzing relevant data, defining adjustment multipliers, and calibrating and validating the model using real-world case studies.

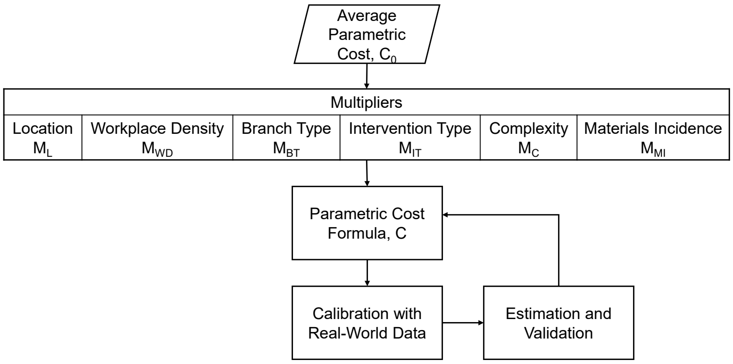

Fig. 1. The estimation framework.

The Average Parametric Cost (based on floor surface) was multiplied by six factors (multipliers) associated to each key parameter:

- Localization multiplier (M_L) related to the geographic localization of the real estate.
- Workplace Density multiplier (M_{WD}), related to the number of required workplaces.
- Branch Type multiplier (M_{BT}), related to material quality and technological equipment (Flagship/Standard).
- Intervention Type multiplier (M_{IT}), new construction or renovation (different categories).
- Complexity multiplier (M_C), related to the complexity of the renovation activities due to the age of the building (specifically historical buildings) or to buildings with special context factors.
- Incidence of materials cost multiplier (M_{MI}).

Labor and material costs were extracted by regional labor cost data (provinces of Milan, Ancona, Naples) and material price indices from ISTAT and the Chamber of Commerce of Milan. Cost of materials were obtained according to Italian public procurement regulations (Legislative Decree n. 50/2016), regional price lists for the relevant time periods, and OMI (Real Estate Market Observatory) for real estate market trends. Input from construction management professionals and banking sector experts were used to refine model parameters and to validate assumptions.

3.3. Analysis of the Cost Multipliers

All the multipliers were determined by statistical analyses of data from ISTAT and National/Regional Observatory of Construction costs.

Geographical characterization (Localization) is fundamental, as it distinguishes areas of the country with different cost influences, specifically labor costs, and materials costs. Cost differences are reported in Table 1, referring to the labor market in large cities, where labor costs are generally higher.

Table 1. Labor cost differences.

Worker's type	Northern Italy (ref. Milan)	Central Italy (ref. Ancona)	Southern Italy (ref. Naples)	m.u.
Common worker (First Level)	23.75	21.59	17.27	€/h
Qualified Worker (Second Level)	26.33	23.94	21.66	€/h
Specialized Worker (Third Level)	28.25	25.68	20.54	€/h
Over Specialized Worker (Forth Level)	29.78	27.07	21.66	€/h
Mean variability from reference value	+10%	Reference value	-20%	-
Localization Multiplier (M_L)	1.1	1	0.8	-

The Workplace Density (see Table 2) is related to the number of workplaces required within the building. This parameter was estimated using linear regression analysis on the basis of the historical dataset (n=183 intervention, ranging from 8 to 45 per year).

The cost of a standard workplace (a "standard workplace" has a floor surface of 11 m2) has remained relatively constant over time, despite the progressive increase in real estate investment. In general, the workplace accounts for 40-45% of total bank space. The cost differences (measured in €/workplace) are limited in the range [-1.75%; +0.44%] in the evaluation period.

Table 2. Workplace Density multiplier (M_{WD}).

Nr. of Workstations	Factor	Sample Size (nr.)	Mean Cost (k€)	SD (k€)	95% CI (k€)	Multiplier
3	Low-density	30	36.0	1.5	[35.45, 36.55]	0.8
5	Standard-density	50	45.0	2.0	[44.43, 45.57]	1.0
7	High-density	20	54.0	2.5	[52.88, 55.12]	1.2

The linear progression of multipliers was derived from observed cost scaling in the historical dataset. The historical data shows that the average cost per workstation is substantially stable over time. Regression analysis indicated a proportional increase in costs with additional workstations within the standard project size range, justifying the use of a linear scale. This approach aligns with industry practice and was validated by the consistency of incremental costs across the dataset; there is no evidence of economies of scale.

For the type of branch only two typologies were considered:

- Flagship (larger floor surface, higher quality level of furniture, finishes, and technological equipment).
- Standard (smaller floor surface, lower quality level of furniture, finishes, and technological equipment).

In Table 3 the Branch Type multipliers are reported.

Table 3. Branch Type multiplier (M_{BT}).

Branch Type	Multiplier
Standard	1.0
Flagship	1.3

Five categories of intervention were considered, as shown in Table 4, due to new construction, transfer, expansion, or extraordinary maintenance. Each type has different costs: the reference level is the medium renovation (Cat. IV).

The concept of age and obsolescence of the building and of its technological dotation is what sometimes strictly determines the complexity of the intervention to be executed. The coefficient of obsolescence considers the building's age and maintenance state. Moreover, historical buildings are protected by the Government and require special considerations and permissions for modifications which impact on project costs. The model considers a specific multiplier for historical buildings, which is determined on a case-by-case basis depending on the complexity and the need for authorization. A scale for the Complexity multiplier (M_C) is reported in Table 5 based on the age of the property.

Table 4. Intervention Type multiplier (M_{IT}).

Category	Type of intervention	Medium cost (€/m^2)	Multiplier
CAT I	New construction	2,047	2.47
CAT II	Heavy renovation (pre-60's aged building)	1,193	1.71
CAT III	Heavy renovation (post- 60's aged building)	1,412	1.44
CAT IV	Medium renovation	827	1.00
CAT V	Light Renovation	407	0.49

Table 5. Complexity multiplier (M_C).

Building Age (minimum)	Multiplier
1	0.91
5	0.95
10	1.00
15	1.05
20	1.075
25	1.10
30	1.125
40	1.175

The analysis considers the Incidence of materials cost like concrete, steel for reinforced concrete, drywall, and marble floors, which have shown a general increase over the years, with an estimated trend of +10% in the analyzed period.

3.4. Parametric Cost Formula

As shown in Fig. 1, in the first formulation the Parametric Cost Formula is based on the multiplication of the Average Parametric Cost (C_0) and of the defined Multipliers. The total estimated cost is calculated as follows:

- Localization multiplier (M_L) related to the geographic localization of the real estate.
- Workplace Density multiplier (M_{WD}), related to the number of required workplaces.
- Branch Type multiplier (M_{BT}), related to material quality and technological equipment (Flagship/Standard).
- Intervention Type multiplier (M_{IT}), new construction or renovation (various categories).
- Complexity multiplier (M_C), related to the complexity of the renovation activities due to the age of the building (specifically historical buildings) or to buildings with special context factors.
- Incidence of materials cost multiplier (M_{MI}).

To consider a scaling factor required by experts a scaling factor formula was used:

$$\text{Scaling factor} = (\text{Project floor surface} / \text{Reference floor surface})^{\alpha-1} \quad (1)$$

Where the Reference floor surface value is 150 m^2 and the scale factor α was assumed as 0.5. According to these modifications the formula for the Parametric Costs is reported in (2):

$$C0 = \text{Average parametric cost} \times \text{Scaling factor} \times ML \times MWD \times MBT \times MIT \times MC \times MMI \quad (2)$$

4. Model validation and discussion.

The model was validated by considering five different real estate management cases. The cases were selected from a larger pool of bank branch projects to ensure diversity in location, intervention type, and project scale. Selection criteria included data completeness, representativeness of typical scenarios, and availability of post-completion cost data. For each case, identified as in Table 6, the main characteristics are reported with the estimated cost, the actual cost, and the variance (in the last three columns).

The results showed that in this formulation the model is quite reliable for standard-sized surfaces (better under 230 m²), with deviations between -1% and +7% compared to actual costs; but, for larger properties (over 250 m²), accuracy gradually decreases and the model overestimated costs by 34%, indicating the need for other correction multipliers for properties outside the standard size range. A deeper analysis showed that two scaling factors would be more efficient: in formula (2) it is suggested to use a scaling factor of 0.5 for standard sized properties (under 230 m²) and a scaling factor of 0.1 for larger properties (over 250 m²). In this way, the error of Estimating and Actual costs for larger properties is limited within the range ±6% as shown in Table 7.

Table 6. Model validation.

Project	Site	Age (yr)	CAT	Type of branch	Nr of workstation	Surface dimension (m^2)	Estimated cost (€/m^2)	Actual cost (€/m^2)	Variance
FON	South	21	V	Std	3	94	707.10	656	+7%
P01	South	21	IV	Std	7	150	1,388.31	1409	-1%
P15	South	10	III	Std	9	230	1,667.91	1640	+5%
MOS	Central	21	V	Std	9	247	724.73	628	+13%
MO2	Central	21	IV	Std	6	375	925.17	612	+34%

Table 7. Modified Model results.

Project	Site	Age (yr)	CAT	Type of branch	Nr of workstation	Surface dimension (m^2)	Estimated cost (€/m^2)	Actual cost (€/m^2)	Variance
FON	South	21	V	Std	3	94	707.10	656	+7%
P01	South	21	IV	Std	7	150	1,388.31	1409	-1%
P15	South	10	III	Std	9	230	1,667.91	1640	+5%
MOS	Central	21	V	Std	9	247	593.66	628	-6%
MO2	Central	21	IV	Std	6	375	641.28	612	+5%

The small validation sample (n=5) limits the statistical power of the results: the variance parameter is reported to reflect this uncertainty.

It is noteworthy that, at the moment, the model does not consider risks and contingency reserves. Additionally, while the model adjusts for historical trends in material costs, it may not fully capture rapid or unexpected market fluctuations. The complexity multiplier for historical buildings also introduces a degree of subjectivity that could affect consistency.

5. Conclusions

This research presented a hybrid parametric methodology for construction cost estimation, focusing on the banking sector. By integrating historical cost data, project-specific parameters, and contextual adjustment multipliers, the model addresses accuracy, adaptability, and transparency challenges found in both model-driven and data-driven approaches. Validation through real-world bank branch projects demonstrates high reliability for standard-sized properties. For larger properties, predictive accuracy improved significantly by applying a modified scaling factor, reducing initial overestimation. The model's interpretable multipliers and sector-specific calibration support early-stage budgeting and resource allocation, enabling informed decisions even with limited project information—an advantage over non-interpretable machine learning models. Its hybrid nature bridges expert-driven and data-centric methods, and its design ensures scalability for other standardized construction contexts.

Despite these strengths, some limitations persist. The model's accuracy decreases for non-standard-sized properties without further adjustment, and the model does not yet include risk or contingency reserves, which are vital for comprehensive forecasting. Future research should extend calibration to other sectors, integrate risk and contingency factors, and enhance responsiveness to market dynamics. Implementing k-fold cross-validation and scenario-based testing will further strengthen predictive robustness, especially under volatile market conditions. Digital integration with BIM or cloud-based platforms offers promising avenues for real-time data sharing and improved project stakeholders collaboration.

References

[1] Tayefeh Hashemi, S., Ebadati, O.M. & Kaur, H. Cost estimation and prediction in construction projects: a systematic review on machine learning techniques. *SN Appl. Sci.* **2**, 1703 (2020). https://doi.org/10.1007/s42452-020-03497-1.

[2] Project Management Institute. (2021). A Guide to the Project Management Body of Knowledge (PMBOK® Guide) (7th ed.).

[3] NASA Cost Estimating Handbook Version 4.0 (2024) Appendix C – Cost Estimating Methodologies (Updated 2/2015).

[4] İnan, T., Narbaev, T., and Hazir, Ö. (2022) A machine learning study to enhance project cost forecasting. IFAC PapersOnLine, 55 (10), pp. 3286-3291.

[5] Narbaev, T., Hazir, Ö., Khamitova, B, and Talgat, S. (2024) A machine learning study to improve the reliability of project cost estimates, International Journal of Production Research, 62:12, 4372-4388.

[6] Akanbi, T., and Zhang, J. (2021) Design information extraction from construction specifications to support cost estimation. Automation in Construction, 131 (2021), Article 103835, 10.1016/j.autcon.2021.103835

[7] ForouzeshNejad, A. A., Arabikhan, F., & Aheleroff, S. (2024). Optimizing Project Time and Cost Prediction Using a Hybrid XGBoost and Simulated Annealing Algorithm. Machines, 12(12), 867. https://doi.org/10.3390/machines12120867

[8] Narbaev, T., Hazir, Khamitova, B., & Talgat, S. (2024). A Machine Learning Study to Improve the Reliability of Project Cost Estimates. International Journal of Production Research, 62(12), 4372–4388, doi.org/10.1080/00207543.2023.2262051

[9] Lathong, K., and Wisaeng, K. (2024). An Innovative Hybrid Machine Learning Techniques For Predicting Construction Cost Estimates. International Journal for Computational Civil and Structural Engineering, 20(3), 69-83. https://doi.org/10.22337/2587-9618-2024-20-3-69-83.

[10] Elmousalami, H. (2021). Machine learning for conceptual cost estimation: Review, implementation and analysis. https://www.sciencedirect.com/science/article/pii/S2352711021000102].

[11] Owolabi, H., et al. (2020). Big data predictive analytics for risk analysis in public-private partnership projects. https://www.mdpi.com/2071-1050/12/3/1042].

[12] Zhang, X. (2025) Dynamic Cost Estimation and Optimization Strategy in Engineering Cost Combining Reinforcement Learning. Applied Mathematics and Nonlinear Sciences, 10(1), 1-14.

[13] Kolyshkina, I., and Simoff S (2021) Interpretability of Machine Learning Solutions in Public Healthcare: The CRISP-ML Approach. Front. Big Data 4:660206. doi: 10.3389/fdata.2021.660206.

[14] Deepa, G., Niranjana, A.J., and Balu, A.S. (2025) A hybrid machine learning approach for early cost estimation of pile foundations. Journal of Engineering, Design and Technology, 23 (1): 306–322. https://doi.org/10.1108/JEDT-03-2023-0097.

[15] Wang, R., Asghari, V., Cheung, C.M., Hsu, S.C., and Lee, C.J. (2022). Assessing effects of economic factors on construction cost estimation using deep neural networks. Automation in Construction, 134, 104080. https://doi.org/10.1016/j.autcon.2021.104080.

[16] Chakraborty, D., Elhegazy, H., Elzarka, H., and Gutierrez, L. (2020) A novel construction cost prediction model using hybrid natural and light gradient boosting. Advanced Engineering Informatics, 46, 10.1016/j.aei.2020.101201.

[17] Pham, V.K., Nguyen, N.A., Vu, T.N.Q., and Phan, T.T. (2025). Impact of Public, Private, and Foreign Direct Investments on Provincial Economic Growth: Evidence From a Transition Country. International Journal of Asian Business and Information Management, 16(1). doi.org/10.4018/IJABIM.366309.

[18] Velumani, P., and Nampoothiri, N.V.N. (2019) Analysis of construction cost prediction studies–global perspective International Review of Applied Sciences and Engineering 10(3) 275-281.

[19] Mosly, I. (2024). Construction Cost-Influencing Factors: Insights from a Survey of Engineers in Saudi Arabia. Buildings, 14(11), 3399. https://doi.org/10.3390/buildings14113399.