RGC Ref. No.:
UGC/FDS25/E06/20
(please insert ref. above)

RESEARCH GRANTS COUNCIL
COMPETITIVE RESEARCH FUNDING SCHEMES FOR THE LOCAL SELF-FINANCING DEGREE SECTOR

FACULTY DEVELOPMENT SCHEME (FDS)

Completion Report
(for completed projects only)

Submission Deadlines:
1. *Auditor's report with unspent balance, if any: within **six** months of the approved project completion date.*
2. *Completion report: within **12** months of the approved project completion date.*

Part A: The Project and Investigator(s)

1. Project Title

Modelling building rehabilitation costs to combat the problems of ageing apartment buildings in Hong Kong

2. Investigator(s) and Academic Department(s) / Unit(s) Involved

Research Team	Name / Post	Unit / Department / Institution
Principal Investigator	LAU Wai-kin / Assistant Professor	Department of Design & Architecture / THEi
Co-Investigator(s)	YAU Yung/Professor of Urban Studies	Department of Sociology & Social Policy/Lingnan University
Co-Investigator(s)	HO Daniel Chi-wing (Retired)	Faculty of Design & Environment / THEi
Co-Investigator(s)	CHAN Hon-chuen (Retired)	Faculty of Design & Environment / THEi
Co-Investigator(s)	TSE Tsz-chun / Assistant Professor	Department of Design & Architecture / THEi
Others	-	-

3. Project Duration

	Original	Revised	Date of RGC / Institution Approval *(must be quoted)*

Project Start Date	1 January 2021	-	-
Project Completion Date	31 December 2022	30 June 2023	1 December 2022 (Institutional Approval)
Duration *(in month)*	24	30	1 December 2022 (Institutional Approval)
Deadline for Submission of Completion Report	31 December 2023	30 June 2024	1 December 2022 (Institutional Approval)

4.3 Please attach photo(s) f acknowledgement of RGC-funded facilities / equipment.

NA

Part B: The Final Report

5. Project Objectives

5.1 Objectives as per original application

1. *To develop a set of measures and indicators for measuring building rehabilitation projects in general and the project costs in particular;*
2. *To identify the significant factors that affect building rehabilitation costs in Hong Kong and overseas;*
3. *To apply machine learning techniques to develop models for explaining and predicting building rehabilitation costs in Hong Kong's apartment buildings;*
4. *To explore the relationships between building rehabilitation costs, and endogenous and exogenous socioeconomic factors; and*
5. *To detect bid-rigging cartels using machine learning techniques and theorise bid-rigging cartels in building rehabilitation projects.*

5.2 Revised objectives

Date of approval from the RGC: NA

Reasons for the change:

1.

2.

3. ..

5.3 Realisation of the objectives
(Maximum 1 page; please state how and to what extent the project objectives have been achieved; give reasons for under-achievements and outline attempts to overcome problems, if any)

For objectives 1 to 2, a comprehensive literature review and structured interviews were conducted to gain insights and validate the findings from the former. The potential problems of presenting building rehabilitation costs per unit and the use of cost indicators were highlighted. The review covered measures and indicators used in construction management research to evaluate project success and performance, as well as factors affecting and/or predictors of construction or maintenance costs. Identified measures and indicators include project scale and characteristics (e.g. gross floor area, external wall area and elemental distribution of costs). Similarly, factors affecting maintenance costs and predictors of construction costs were referenced and identified. Part of the findings from these two objectives were presented at the AUBEA 2022 conference in Nov 2022, in Penrith, Australia. The remaining findings will be disseminated in journal publications, and the objectives are considered to have been substantially achieved.

For objective 3, machine learning techniques were shortlisted before developing the models. Traditional supervised learning methods such as Support Vector Machine (SVM), Decision Tree (DT) and Random Forest (RF) and contemporary supervised learning methods such as Deep Neural Network (DNN) were selected. Based on the findings from objectives 1 and 2, model parameters were determined. For example, gross floor area, number of units, number of storeys and scope of work were introduced into the models. The dataset size was 259 and 233 after removing outliers. Despite repeated trials, the developed models did not show satisfactory predictive power. The research team is still looking for ways to generate more reliable models. Nonetheless, this objective is considered to have been substantially achieved, and a working paper titled “”Machine learning approach for predicting the rehabilitation cost of high-rise residential buildings in Hong Kong” has been prepared.

For objective 4, the research team explored the association between socio-economic characteristics of the residents and building rehabilitation costs. Median income, education level, ownership and property value (which was later removed) were tested for the association with building rehabilitation cost using correlation analysis and multiple linear regression. For this purpose, models making use of Bayesian Networks, XGBoost, Support Vector Regression (SVR), Quadratic regression and Lasso regression were developed. So far, there is no strong evidence suggesting a causal effect of socio-economic factors on rehabilitation costs, except for ownership. A working paper titled “Detecting Price Discrimination Using Frequentist and Bayesian Multiple Linear Regression” was drafted while the research team looks for ways to enhance the results. That said, this objective has been substantially achieved.

For objective 5, the research team sought to understand several aspects related to bid-rigging in the construction industry, such as the reasons of existence of bid-rigging in the construction industry despite measures such as prequalification, the methods of detecting bid-rigging (whether using machine learning approaches or not) and the schemes by the government or others to prevent bid-rigging. A working paper titled “Machine-learning based bid rigging detection of building rehabilitation sector in Hong Kong: Conceptual model” was drafted and is being reviewed and fine-tuned by the team before dissemination. The background, forms of bid-rigging, factors promoting bid-rigging, and screens of bid-rigging are discussed. The framework of the machine learning detection model was also proposed. This objective is considered to have been substantially completed.

5.4 Summary of objectives addressed to date

Objectives *(as per 5.1/5.2 above)*	**Addressed** *(please tick)*	**Percentage Achieved** *(please estimate)*
1. *To develop a set of measures and indicators for measuring building rehabilitation projects in general and the project costs in particular*	Yes	100%
2. *To identify the significant factors that affect building rehabilitation costs in Hong Kong and overseas*	Yes	100%
3. *To apply machine learning techniques to develop models for explaining and predicting building rehabilitation costs in Hong Kong's apartment buildings*	Yes	100%
4. *To explore the relationships between building rehabilitation costs, and endogenous and exogenous socioeconomic factors*	Yes	100%
5. *To detect bid-rigging cartels using machine learning techniques and theorise bid-rigging cartels in building rehabilitation projects*	Yes	100%

6. Research Outcome

6.1 Major findings and research outcome
(Maximum 1 page; please make reference to Part C where necessary)

Measures and indicators in construction management research and common unit rates developed by the Urban Renewal Authority (URA) were considered. It was concluded that that it is not necessary to develop another set of common unit rates. It is better to use them to improve cost transparency and track cost trends of building rehabilitation projects. Besides, suitable measures and indicators for building rehabilitation projects, such as building rehabilitation cost per m^2 and elemental distribution of cost, were identified and considered appropriate. Project scale and characteristic measures/indicators not only serve as measures but were also identified as predictors of building rehabilitation costs, e.g. site area, building height, building age and number of flats. Together with other identified factors such as the duration of rehabilitation project, they were used as independent variables for developing the prediction models.

As mentioned in 5.3, DNN, SVM, DT, RFR, etc. were used to develop prediction models. The input variables are gross floor area, number of units, number of storeys and building age. Trials using other input variables were also conducted. The size of the dataset is 259 and 233 after removing outliers, giving the following results (R^2, adjusted R^2 and RMSE):

DNN: -2.133, -2.188, 2,148,224
SVM: -0.03, -0.13, 1,377,804
DT: - , - , 1.003×10^{12}
RFR: 0.8, - , 3.2×10^{11}
XGBoost: 0.85, 0.83, 527,645
Quadratic regression: 0.81, 0.80, 524585

For the correlation between socioeconomic characteristics of the residents and building rehabilitation costs, median income level, education level, ownership and property value have been tested with building rehabilitation costs. Later property value was not considered. The results from Ordinary Least Square (OLS) multiple linear regression (MLR) model is presented below:

OLS MLR	Intercept: 0.056, + and insignificant
R^2 & adjusted R^2: 78.1% & 77.6%	Median Income (MI): -0.026, - and insignificant
RMSE: 0.108	Gross Floor Area (GFA): 0.745, + and significant
MAD: 0.078	No. of Storeys (S): 0.117, + and significant
	Building Age (A): -0.04, - and insignificant
	Ownership (O): 0.086, + and significant

* Education Level (E) and No. of Units (U) removed due to multi-collinearity

As can be seen above, only ownership factor has a substantial causal effect on rehabilitation cost. In addition, Bayesian Networks, XGBoost, SVR, Quadratic regression and Lasso regression were developed but none of these of these models indicate sufficient predictive power nor meaningful casual effects.

The major findings will be disseminated through journal publications and several working papers have been prepared. As reiterated in this report, the research team is working on improving the results of the models. In Part C, the abstracts of the journal publications that are under preparation are detailed in Appendices 1 to 3.

6.2 Potential for further development of the research and the proposed course of action *(Maximum half a page)*

It has been confirmed in this and previous studies by the PI that the ageing of building stock is occurring in Hong Kong. The study not only provided insights but also enabled the monitoring of building rehabilitation cost trends over the long term. With a better understanding of the factors affecting building rehabilitation costs and the use of machine learning prediction models, it becomes easier and more convenient to predict building rehabilitation costs in advance.

Society as a whole has increasingly acknowledged the imminence of ageing building stock and the necessity of building maintenance and rehabilitation. Initiatives such as preventive maintenance, maintenance funds and insuring buildings are now on the table. While we may not be able to lower building rehabilitation costs, we can explain the reasons behind high rehabilitation costs and determine the costs more objectively.

Looking forward, collaboration with bodies such as the Urban Renewal Authority will follow to enhance and test the prediction models using new empirical data. As reported in the other sections of this report, findings and insights from this study have been and will be shared with students studying surveying and related subjects. The knowledge generated in this study paves the way for further research into building maintainability. The subject of building maintainability has been under-studied, except for a series of endeavours in Singapore a decade ago aimed at enhancing building maintainability. In other engineering disciplines, maintainability and reliability have been more extensively studied. This study has sown the seeds for further research on building maintainability. For example, if site constraints and inadequate access push up rehabilitation costs, building maintainability can be improved by alleviating these physical and managerial constraints and by providing more adequate access to facilitate building maintenance and rehabilitation. It is believed that this project has laid the foundation of many subsequent, in-depth studies, such as benchmarking the time and cost performance of building rehabilitation projects and for more policy discussions.

7. Layman's Summary

(Describe <u>in layman's language</u> the nature, significance and value of the research project, in no more than 200 words)

In this study, building rehabilitation cost prediction using machine learning techniques was investigated. A comprehensive set of measures and indicators suitable for building rehabilitation projects was examined, and key factors affecting these costs were identified. A challenge in using these measures is balancing simplicity with the risk of oversimplification, as the scope can vary significantly across projects. Various techniques in developing machine learning models for cost prediction were considered. Scope-related parameters can be included only in the Random Forest (RF) Model, but not in Deep Neural Network (DNN), Support Vector Machine (SVM) and Decision Tree (DT) models due to their binary nature. All models are undergoing fine-tuning to provide higher predictive power. For the association between socio-economic factors and building rehabilitation costs, results indicate that only ownership has a substantial causal effect. There is no significant impact or positive relationship with rehabilitation costs, and thus price discrimination and willingness to pay cannot be substantiated. For bid-rigging detection, a review was conducted to explored its persistence, detection methods, countries most known for bid-rigging, actions taken by those countries, screens suggested by researches and gaps in Hong Kong.

Part C: Research Output

8. Peer-Reviewed Journal Publication(s) Arising Directly From This Research Project

(Please attach a copy of the publication and/or the letter of acceptance if not yet submitted in the previous progress report(s). All listed publications must acknowledge RGC's funding support by quoting the specific grant reference.)

The Latest Status of Publications				**Author(s)** *(denote the corresponding author with an asterisk*)*	**Title and Journal / Book** *(with the volume, pages and other necessary publishing details specified)*	**Submitted to RGC** *(indicate the year ending of the relevant progress report)*	**Attached to this Report** *(Yes or No)*	**Acknowledged the Support of RGC** *(Yes or No)*	**Accessible from the Institutional Repository** *(Yes or No)*
Year of Publication	**Year of Acceptance** *(For paper accepted but not yet published)*	**Under Review**	**Under Preparation** *(optional)*						
N/A	N/A	N/A	√	Lau, W. K.*, Sumanarathna, N. & Yau, Y.	Machine learning approach for predicting the rehabilitation cost of high-rise residential buildings in Hong Kong	No	Yes	Yes	No
N/A	N/A	N/A	√	Lau, W. K.*, Sumanarathna, N. & Yau, Y.	Detecting Price Discrimination Using Frequentist and Bayesian Multiple Linear Regression	No	Yes	Yes	No
N/A	N/A	N/A	√	Lau, W. K.*, Sumanarathna, N. & Yau, Y.	Machine-learning based bid rigging detection of building rehabilitation sector in Hong Kong: Conceptual model"	No	Yes	Yes	No

The abstracts of these three manuscripts are attached as Appendices 1 to 3.

9. Recognized International Conference(s) In Which Paper(s) Related To This Research Project Was / Were Delivered
(Please attach a copy of each conference abstract)

Month / Year / Place	**Title**	**Conference Name**	**Submitted to RGC** *(indicate the year ending of the relevant progress report)*	**Attached to this Report** *(Yes or No)*	**Acknowledged the Support of RGC** *(Yes or No)*	**Accessible from the Institutional Repository** *(Yes or No)*
Nov 2022, Australia	Developing Machine Learning Models for Building Rehabilitation Cost Prediction	AUBEA 2022 – The 45th Australasian Universities Building Education Association Conference	No	Yes	Yes	No

The abstract submitted to the conference is attached as Appendix 4.

10. Whether Research Experience And New Knowledge Has Been Transferred / Has Contributed To Teaching And Learning
(Please elaborate)

- Affirmative. The PI has transferred the research skills and insights gained from this project to the following learning modules: DSU5307 Graduation Project 1: Research Proposal and DSU5406 Graduation Project 2: Thesis Report when supervising students' graduation project Work. As reported below, one student under the PI's supervision studied machine learning in her graduation projects. For others taking DSU5307 and DSU5406 modules, the PI shared machine learning techniques and research experience from this project with them when talk about research methodology and proposal writing (roughly 80 to 100 students in 3 years).
- A research seminar titled "Developing Machine Learning Models for Building Rehabilitation Cost Prediction" was held on 21 March 2024 to share the project findings to THEi and VTC community.

11. Student(s) Trained

(Please attach a copy of the title page of the thesis)

Name	Degree Registered for	Date of Registration	Date of Thesis Submission / Graduation
[redacted]	Surveying	Sep 2018	Jun 2022

12. Other Impact

(e.g. award of patents or prizes, collaboration with other research institutions, technology transfer, teaching enhancement, etc.)

The knowledge gained in the research process and findings were also transferred to enhance teaching and learning in various modules of the BSc (Hons) in Surveying relating to building maintenance, facilities management and conversion of buildings.

13. Statistics on Research Outputs

	Peer-reviewed Journal Publications	Conference Papers	Scholarly Books, Monographs and Chapters	Patents Awarded	Other Research Outputs (please specify)	
No. of outputs arising directly from this research project	0	1	0	0	Type	No.
					N/A	0

14. Public Access Of Completion Report

(Please specify the information, if any, that cannot be provided for public access and give the reasons.)

Information that Cannot Be Provided for Public Access	Reasons
NA	