RGC Ref. No.:
UGC/FDS14/P05/17
(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

Statistical disclosure control of confidential data using vine copulas with applications

to finance and healthcare management

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

Research Team	Name / Post	Unit / Department / Institution
Principal Investigator	Dr. LAM Shu Yan Associate Professor	Department of Mathematics, Statistics and Insurance/The Hang Seng University of Hong Kong
Co-Investigator	Dr. CHU Man Ying Assistant Professor	Department of Social Sciences/The Education University of Hong Kong
Co-Investigator	Prof. SO Mike K P Professor	Department of Information Systems, Business Statistics & Operations Management/The Hong Kong University of Science and Technology
Co-Investigator	Prof. TIWARI Agnes Honorary Professor	School of Nursing/The University of Hong Kong

3. Project Duration

	Original	Revised	Date of RGC / Institution Approval (must be quoted)
Project Start Date	01/01/2018		
Project Completion Date	31/12/2020	30/06/2021	17th, Aug, 2020
Duration (in month)	36	42	17th, Aug, 2020
Deadline for Submission of Completion Report	31/12/2021	30/06/2022	17th, Aug, 2020

Part B: The Final Report

5.   Project Objectives

5.1   Objectives as per original application

1.   To develop a vine copula approach for statistical disclosure control.
2.   To study theoretical properties with respect to the utility and risk of the proposed vine copula approach.
3.   To perform comprehensive experiments on the effectiveness of using vine copulas for data perturbation. Specifically, the study will examine how well the disclosure risk control and the nonlinear relationships between confidential and non-confidential attributes are preserved.
4.   To apply and evaluate the proposed vine copula approach to finance and healthcare management using real data.
5.   To extend the approach to handle mixed-type data.

5.2   Revised objectives

Date of approval from the RGC: N/A

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)

With close collaboration and great effort paid by the team members, we have completed the project with two papers published and one paper under revision. The project objectives have been fully achieved.

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 vine copula approach for statistical disclosure control.	√	100%
2. To study theoretical properties with respect to the utility and risk of the proposed vine copula approach.	√	100%
3. To perform comprehensive experiments on the effectiveness of using vine copulas for data perturbation. Specifically, the study will examine how well the disclosure risk control and the nonlinear relationships between confidential and non-confidential attributes are preserved.	√	100%
4. To apply and evaluate the proposed vine copula approach to finance and healthcare management using real data.	√	100%
5. To extend the approach to handle mixed-type data.	√	100%

6. Research Outcome

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

In the paper published in 2019, we introduced and demonstrated the use of two useful statistical disclosure control (SDC) methods, a general additive perturbation method and the copula perturbation method, in an empirical study. The results from our empirical study show that using the general additive perturbation method and the copula perturbation method help to release perturbed data of their confidential attributes and protect the privacy of respondents. At the same time, these methods retain the statistical information of the original database. This allows the perturbed database to be used for

further research and analysis without the need of having access to the original database and therefore the opportunities for the research and analysis of interesting but sensitive topics in business, social sciences and health care will be feasible.

In the paper published in 2022, we extended the skew-t data perturbation (STDP) to develop a new SDC method for data with continuous variables. In this new SDC method, we constructed an extended skew-t (EST) copula to release confidential data for third-party usage. Using the EST copula for producing perturbed data, we can incorporate rich statistical information in the perturbed data while preserving the marginal distributions of the data. An advancement of this EST-SDC method is to use a copula distribution, which allows generation of perturbed data from bivariate conditional EST copulas sequentially. Simulations and a real data study were included to demonstrate how the EST-SDC method can be applied and to compare with the STDP method.

In the under-revision paper, we developed a new SDC method for mixed-type data based on vine copula. Our vine-SDC method improves the SDC method using the extended skew-t copula by allowing the bivariate copulas in the vine decomposition to take various forms, thus offering a better fit for the joint distribution of the data and more flexibility in data perturbation. A simulation and a study of real survey data were used to explore the performance and strength of vine-SDC.

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

There is a great potential for further development of the research as the methodology can be used to protect confidential information which is an important topic in social sciences, medical and business research.

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 project, we developed a vine copula approach that aims to capture and preserve nonlinear patterns in confidential and non-confidential data after data perturbation. The use of the vine copula approach enhances the utility of the perturbed data.

In terms of theoretical and methodological contributions, the project advances the statistical disclosure control method and improve the computational efficiency. To show the practical value of the proposed vine copula approach, we conducted a series of numerical experiments to investigate the performance of the proposed approach and gain insights into the implementation. Applications to cases in business and healthcare management were studied to showcase the advantages of the use of vine copulas. The theoretical and practical findings of this study are expected to provide a pathway and guidelines on statistical disclosure control for more comprehensive data analysis based on nonlinear patterns in perturbed data.

The collaboration among the PI and Co-Is also helps to promote interdisciplinary and cross-departmental research among local self-financing degree-awarding institutions and universities.

Part C: Research Output

8.   **Peer-Reviewed Journal Publication(s) Arising <u>Directly</u> 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)*						
2019				Chu, A. M. Y., Lam, B. S. Y., Tiwari, A., & So, M, K. P.*	An empirical study of applying statistical disclosure control methods to public health research. *International al Journal of Environmental Research and Public Health*, 16(22), 4519. https://doi.org/10.3390/ijerph16224519	No	Yes (Annex I)	Yes	Yes (https://researchdb hsu.edu hk/view/publication/201900296)
2021				Chu, A. M. Y., Ip, C. Y., Lam B. S. Y., & So, M. K. P.*	Statistical disclosure control for continuous variables using an extended skew-t copula. *Applied Stochastic Models in Business and Industry*, 38(1), 96-115. https://doi.org/10.1002/asmb.2650	No	Yes (Annex II)	Yes	Yes (https://researchdb hsu.edu hk/view/publication/202100213)

	2022 (under revision)			Chu, A. M. Y., Ip, C. Y., Lam, B. S. Y., & So, M. K. P.*	Vine Copula Statistical Disclosure Control for Mixed-type Data, *Computational Statistics & Data Analysis*	No	Yes (abstract) (Annex III)	Yes	No

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)*
07/2019 Japan	Statistical disclosure control of confidential data with an application to social research	The 5th International Conference on Education, Psychology, and Social Studies	2019	No	Yes (in the talk)	Yes (https://researchdb.hsu.edu.hk/view/publication/201900372)
01/2020 Japan	Applying statistical disclosure control in business research	2020 International Conference on Business, Information, Tourism, and Economics	No	Yes (Annex IV)	Yes (in the talk)	Yes (https://researchdb.hsu.edu.hk/view/publication/202000252)

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

    Yes. For example, a co-author included the concept of statistical disclosure control in one of his teaching courses.

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
N/A			

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

N/A

13. Statistics on Research Outputs

	Peer-reviewed Journal Publications	Conference Papers	Scholarly Books, Monographs and Chapters	Patents Awarded	Other Research Outputs (please specify)	
					Type	No.
No. of outputs arising directly from this research project	2	2				

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
N/A	