RGC Ref. No.:
UGC/FDS14/P01/16
(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

Advanced statistical methods for complex longitudinal data analysis

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

Research Team	Name / Post	Unit / Department / Institution
Principal Investigator	TANG Man-Lai/Professor	Mathematics, Statistics and Insurance/The Hang Seng University of Hong Kong
Co-Investigator(s)	Nil	Nil
Others	Nil	Nil

3. Project Duration

	Original	Revised	Date of RGC / Institution Approval *(must be quoted)*
Project Start Date	01/01/2017		
Project Completion Date	31/12/2019		
Duration *(in month)*	36		
Deadline for Submission of Completion Report	31/12/2020		

Part B: The Final Report

5. Project Objectives

5.1 Objectives as per original application

1. (**Non-normal longitudinal data analysis**) A profile likelihood estimation approach, in which the estimated likelihood is maximized without the specification of the working correlation structure, is proposed for non-normal longitudinal data.
2. (**Response with fixed detection limits and measurement error in covariates**) A censored quantile regression (QR) with mixed effects and covariates in errors is proposed. Algorithm that combines inverse probability censoring weighted (IPCW) and orthogonal regression (OR) methods will be developed for parameter estimation.
3. (**Asymptotic statistical properties**) Consistency, efficiency and asymptotic normality of the proposed estimates will be proved. Extensive simulation studies will be conducted to evaluate the performance of the proposed methods.
4. Extensive simulation studies will be conducted to evaluate the performance of the proposed methods in Objectives (1) and (2) above.
5. Computer codes prepared in R will be made available to potential users.

5.2 Revised objectives

Date of approval from the RGC: N.A.

Reasons for the change: N.A.

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)

Together with my collaborators, we have successfully developed a profile likelihood estimation approach, in which the estimated likelihood is maximized without the specification of the working correlation structure, for non-normal longitudinal data. The result has been published in *Biometric* (2018, **74**, 220 - 228) (i.e., fulfilled Objectives 1). Besides, we have developed a joint modeling for mixed-effects quantile regression for longitudinal data with detection limits and covariates measured with error. The result has been published in *Computational Statistics* (2018, **33**, 1563 - 1587) (i.e., fulfilled Objectives 2). Asymptotic properties (i.e., efficiency, consistency and unbiasedness etc) have been developed for the proposed estimates in Objectives 1 and 2 (i.e., fulfilled Objective 3) and our simulation studies confirm that our proposed estimates perform satisfactorily (i.e., fulfilled Objective 4). Finally, we have written computer codes (in R) to implement the proposed methodologies and are made available to potential practitioners (i.e., fulfilled Objective 5).

5.4 Summary of objectives addressed to date

Objectives *(as per 5.1/5.2 above)*	**Addressed** *(please tick)*	**Percentage Achieved** *(please estimate)*
1. (**Non-normal longitudinal data analysis**) A profile likelihood estimation approach, in which the estimated likelihood is maximized without the specification of the working correlation structure, is proposed for non-normal longitudinal data.	✔	100%
2. (Response with fixed detection limits and measurement error in covariates) A censored quantile regression (QR) with mixed effects and covariates in errors is proposed. Algorithm that combines inverse probability censoring weighted (IPCW) and orthogonal regression (OR) methods will be developed for parameter estimation.	✔	100%
3. (Asymptotic statistical properties) Consistency, efficiency and asymptotic normality of the proposed estimates will be proved. Extensive simulation studies will be conducted to evaluate the performance of the proposed methods.	✔	100%
4. Extensive simulation studies will be conducted to evaluate the performance of the proposed methods in Objectives (l) and (2) above.	✔	100%
5. Computer codes prepared in R will be made available to potential users.	✔	100%

6. Research Outcome

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

Inappropriate choice of working correlation structure in generalized estimating equations (GEE) could lead to inefficient parameter estimation while impractical normality assumption in likelihood approach would limit its applicability in longitudinal data analysis. In this article, we propose a profile likelihood method for estimating parameters in longitudinal data analysis via maximizing the estimated likelihood. The proposed method yields consistent and efficient estimates without specifications of the working correlation structure nor the underlying error distribution. Both theoretical and simulation results confirm satisfactory performance of the proposed method. We illustrate our methodology with a diastolic blood pressure data set. These results have been published in *Biometric* (2018, **74**, 220 - 228).

It is very common in AIDS studies that response variable (e.g., HIV viral load) may be subject to censoring due to detection limits while covariates (e.g., CD4 cell count) may be measured with error. Failure to take censoring in response variable and measurement errors in covariates into account may introduce substantial bias in estimation and thus lead to unreliable inference. Moreover, with non-normal and/or heteroskedastic data, traditional mean regression models are not robust to tail reactions. In this case, one may find it attractive to estimate extreme causal relationship of covariates to a dependent variable, which can be suitably studied in quantile regression framework. In this paper, we consider joint inference of mixed-effects quantile regression model with right-censored responses and errors in covariates. The inverse censoring probability weighted method and the orthogonal regression method are combined to reduce the biases of estimation caused by censored data and measurement errors. Under some regularity conditions, the consistence and asymptotic normality of estimators are derived. Finally, some simulation studies are implemented and a HIV/AIDS clinical data set is analyzed to illustrate the proposed procedure. The result has been published in *Computational Statistics* (2018, **33**, 1563 - 1587).

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

Based on the results of profile likelihood method, we further developed a so-called profile likelihood ratio test (PLRT) based on the estimated error density for the multiple linear regression model. Unlike the existing likelihood ratio test (LRT), our proposed PLRT does not require any specification on the error distribution. The asymptotic properties are developed and the Wilks phenomenon is studied. Simulation studies are conducted to examine the performance of the PLRT. It is observed that our proposed PLRT generally outperforms the existing LRT, empirical likelihood ratio test and the weighted profile likelihood ratio test in sense that (i) its type I error rates are closer to the prespecified nominal level; (ii) it generally has higher powers; (iii) it performs satisfactorily when moments of the error do not exist (eg, Cauchy distribution); and (iv) it has higher probability of correctly selecting the correct model in the multiple testing problem. A mammalian eye gene expression dataset and a concrete compressive strength dataset are analyzed to illustrate our methodologies. The result was accepted for publication in *Statistics in Medicine* (2020).

Based on the results of quantile regression for mixed effects, we further developed a Bayesian bridge-randomized penalty which is incorporated into the quantile mixed effects models of ordinal longitudinal data to conduct parameter estimation and variable selection simultaneously. The Bayesian joint hierarchical model is established and an efficient Gibbs sampler algorithm is employed to perform posterior statistical inference. Finally, the proposed approach is illustrated using simulation studies and applied to an ordinal longitudinal real dataset of firm bond ratings. The result was accepted for publication in *Computational Statistics* (2020).

7. Layman's Summary

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

Longitudinal data are generally defined as data from observations that are measured repeatedly over a period of time, and they arise frequently in many research areas. A key characteristic of longitudinal data is that observations within the same subject may be correlated, which motivates most of the statistical methods that are used to analyse longitudinal data. Although there have been extensive methodological developments in the analysis of longitudinal data, statistical analyses of complex longitudinal data (e.g., non-normal responses, censored responses and measurement error in covariates) can be very challenging, and advanced statistical methods have great research interest and practical demand.

In this project, a profile likelihood estimation approach, in which the estimated likelihood is maximised without the specification of the working correlation structure, is proposed for non-normal longitudinal data. Also, a censored QR with mixed effects and covariates in errors is proposed. An algorithm that combines IPCW and OR methods will be developed for parameter estimation. Both theoretical and simulation results confirm that our proposed methodologies perform well under various settings.

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 correspond-ing 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)*						
2018				Chen, Ziqi*, **Tang, Man-Lai**, and Gao, Wei	A profile likelihood approach for longitudinal data analysis. *Biometrics* (2018, **74**, 220 – 228).	2018	Yes	Yes	Yes
2018				Tian, Yuzhu, **Tang, Man-Lai***, and Tian, Maozai	Joint modeling for mixed-effects quantile regression of longitudinal data with detection limits ad covariates measured with error, with application to AIDS studies. *Computational Statistics* (2018, **33**, 1563 - 1587).	2018	Yes	Yes	Yes
2018				Tian, Guo-Liang, Liu, Yin, **Tang, Man-Lai***, and Jiang, Xuejun	Type I multivariate zero-truncated/adjusted Poisson distributions with applications. *Journal of Computational and*		Yes	Yes	Yes

					Applied Mathematics (2018, **344**, 132 – 153)				
2019				Xu, Lin, **Tang, Man-Lai***, and Chen, Ziqi	Analysis of longitudinal data by combining multiple dynamic covariate models. *Statistics And Its Interface* (2019, **12**, 479 – 487)		Yes	Yes	Yes
2019				Tian, Guo-Liang, Liu, Yin, **Tang, Man-Lai***, and Li, Tao	A novel MM algorithm and the mode-sharing method in Bayesian computation for the analysis of general incomplete categorical data. *Computational Statistics and Data Analysis* (2019, **140**, 122 – 143)		Yes	Yes	Yes
2019				Liu, Yin, Tian, Guo-Liang, **Tang, Man-Lai***, and Yuen, Kam Chuen	A new multivariate zero-adjusted Poisson model with applications to biomedicine. *Biometrical Journal* (2019, **61**, 1340 – 1370)		Yes	Yes	Yes
2019				Tian, Yuzhu, **Tang, Man-Lai***, Wang, Liyong, and Tian, Maozai	Bayesian bridge-randomized penalized quantile regression estimation for linear regression model with AP(q) perturbation. *Journal of Statistical*		Yes	Yes	Yes

					Computation and Simulation (2019, **89**, 2951 – 2979)				
2019				Qiu, Shi-Fang, Poon, Wai-Yin, **Tang, Man-Lai***, and Tao, Ji-Ran	Construction of confidence intervals for the risk differences in stratified design with correlated bilateral data. *Journal of Biopharmaceutical Statistics* (2019, **29**, 446 – 467)		Yes	Yes	Yes
2020	2020			Yan, Feifei, Xu, Qing-Song, **Tang, Man-Lai***, and Chen, Ziqi	Kernel density-based likelihood ratio tests for linear regression models. *Statistics in Medicine*		Yes	Yes	Yes
2020	2020			Tian, Yuzhu, **Tang, Man-Lai***, Chan, Wai-Sum, and Tian Maozai	Bayesian bridge-randomized penalized quantile regression for ordinal longitudinal data, with application to firm's bond ratings. *Computational Statistics*		Yes	Yes	Yes

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)*
N.A.						

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

No. The results that have been developed are too technical to students in the University.

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)	
No. of outputs arising directly from this research project	10	0	0	0	Type	No.

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.	