UNIVERSITI PUTRA MALAYSIA

GEE-SMOOTHING SPLINE FOR SEMIPARAMETRIC ESTIMATION OF LONGITUDINAL CATEGORICAL DATA

SULIADI

IPM 2011 8

GEE-SMOOTHING SPLINE FOR SEMIPARAMETRIC ESTIMATION OF LONGITUDINAL CATEGORICAL DATA

By

SULIADI

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia in Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics

2011

This thesis is especially dedicated to
my beloved wife - Azkiyah
and my beloved son - Irfan Arif Aprilianto
for the unconditional patient, love and support.

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Doctor of Philosophy

GEE-SMOOTHING SPLINE FOR SEMIPARAMETRIC ESTIMATION OF LONGITUDINAL CATEGORICAL DATA

By

SULIADI

June 2011

Chair: Prof. Noor Akma Ibrahim, PhD

Faculty: Institute for Mathematical Research

In this thesis we propose estimation methods of semiparametric marginal models for longitudinal (correlated) categorical data, where the systematic component of the model consists of parametric and nonparametric forms. We develop GEE-Smoothing spline as a method to analyze semiparametric model for longitudinal data. The proposed methods are an extension of parametric generalized estimating equation (GEE) to semiparametric GEE by introducing smoothing spline into parametric GEE. We derive estimation method of GEE-Smoothing spline in the case of longitudinal binary, ordinal, and nominal data. Derivation of the estimating equation of GEE-Smoothing spline for these three types of categorical data is the same. However their estimating equations have different forms of the covariance and correlation matrices.

In the estimation of the association (correlation) parameter for binary data, we use moment method of Liang & Zeger's and method of Prentice's. For ordinal

and nominal data, we use different models of the covariance matrices than of binary data. These models need smaller number of the association parameter to be estimated which is different from the existing models of parametric GEE for ordinal data. We also derive and propose the methods to estimate the association parameter for these two types of data.

The properties of the estimate for both parametric and nonparametric components of GEE-Smoothing spline are evaluated using simulation studies. We obtained that the estimates of parametric component for binary and ordinal data are unbiased. Whilst for nominal data, the estimates of parametric component are almost unbiased. Meanwhile the estimates of the nonparametric component for all types of data are biased, with the bias decreases when the sample size increases. The estimators of both parametric and nonparametric components are also consistent, and the consistency is not affected by the correct or incorrect working correlation used in model. This consistency property holds for correlated and independent data. The efficiency of the estimates of using independent or correlated working correlation in the estimation depends on the type of covariate, such as time varying, subject specific, or mean-balanced covariates. The estimates of both parametric and nonparametric components also follow the central limit theorem (CLT), for both independent and correlated data, and using correct or incorrect working correlation. Both components estimate have normal distribution.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PELICINAN SPLINE-GEE UNTUK PENGANGGARAN SEMI-BERPARAMETER BAGI DATA LONGITUDINAL BERKATEGORI

Oleh

SULIADI

Jun 2011

Pengerusi: Prof. Noor Akma Ibrahim, PhD

Fakulti: Institut Penyelidikan Matematik

Dalam tesis ini kami mencadangkan kaedah penganggaran model sut semi berparameter bagi data longitudinal (berkorelasi) berkategori, yang mana komponen sistematik bagi model terdiri dari bentuk berparameter dan tak berparameter. Kami bangunkan Pelicinan Spline-GEE sebagai kaedah untuk menganalisis model semi berparameter untuk data longitudinal. Kaedah yang dicadangkan adalah lanjutan dari penganggaran persamaan teritlak (GEE) berparameter kepada semi berparameter GEE dengan memperkenalkan pelicinan spline ke dalam GEE berparameter. Kami huraikan kaedah anggaran Pelicinan Spline-GEE bagi data longitudinal duaan, ordinal, dan nominal. Huraian persamaan anggaran Pelicinan Spline-GEE untuk ketiga jenis data berkategori ini adalah sama. Walau bagaimanapun penganggaran persamaan ini mempunyai bentuk matriks kovarians dan korelasi yang berbeza.

Dalam penganggaran bagi parameter hubungan (korelasi) untuk data duaan, kami menggunakan kaedah momen Liang & Zeger dan kaedah Prentice. Bagi data ordinal dan nominal, kami menggunakan model matriks kovarians yang berbeza dari data duaan. Model-model ini memerlukan bilangan parameter yang kecil untuk dianggar yang berbeza dari GEE berparameter bagi data ordinal yang sedia ada. Kami juga menghuraikan dan mencadangkan kaedah untuk menganggar parameter hubungan bagi kedua-dua jenis data ini.

Sifat-sifat penganggar bagi kedua-dua komponen berparameter dan tak berparameter Pelicinan Spline-GEE dinilai menggunakan kajian simulasi. Komponen berparameter bagi data duaan dan ordinal adalah saksama. Manakala bagi data nominal, penganggar bagi komponen berparameter adalah hampir saksama. Sementara itu penganggar komponen tak berparameter bagi semua jenis data adalah pincang, dengan kepincangan mengecil apabila saiz sampel meningkat. Penganggar komponen berparameter dan tak berparameter adalah konsisten dengan kekonsistenan tidak dipengaruhi oleh struktur korelsi yang betul atau salah yang digunakan dalam model. Sifat konsisten ini berlaku untuk data berkorelasi dan merdeka. Kecekapan penganggar menggunakan sama ada struktur korelasi merdeka atau berkorelasi dalam anggaran bergantung pada jenis kovariat, seperti kovariat yang berubah mengikut masa, kovariat tertentu mengikut subjek, atau kovariat seimbang min. Kedua-dua penganggar berparameter dan tak berparameter juga mengikut teorem had memusat, tidak kira samada data merdeka atau berkorelasi, dengan menggunakan struktur korelasi yang betul atau salah. Penganggar bagi kedua-dua komponen adalah bertaburan normal.

ACKNOWLEDGEMENT

Alhamdulillahirobil'alamin, first of all I would like to express my utmost thanks and gratitude to the Almighty Allah SWT, the Sustainer, the most Gracious and the most Merciful; without whose will no one can achieve anything. My salawat and salam is addressed to His righteous messenger, prophet Muhammad SAW.

I would like to take this opportunity to express my appreciation and gratitude to my supervisory committee, Prof. Dr. Noor Akma Ibrahim, Assoc. Prof. Dr. Isa Daud and Dr. Isthrinayagy S. Krishnarajah for her invaluable guidance, suggestions, discussions, and patience throughout the research and preparation of this thesis.

My sincere gratitude is also extended to the financial support provided by Graduate Research Fellowship (GRF) of Univeristi Putra Malaysia, Science Fund from Ministry of Science, Technology and Innovation Malaysia, and Bandung Islamic University. Acknowledgements are also due to Rector of Bandung Islamic University (UNISBA), Dean of Faculty of Mathematics and Natural Sciences UNISBA, and Head of Dept. of Statistics UNISBA, whose give me an opportunity to pursue my study at Universiti Putra Malaysia.

Special thanks also to all my friends in 685 Jl. 18/2A Taman Sri Serdang, Srikembangan: Bambang SAS, Abdul Kudus, Aris Munawar, Abdul Rahman, Sofjan Hadi, Abdul Malik, Asep Edi K, and all my friends at Indonesian Student Association - UPM (PPI-UPM) for their support and relationship.

Last but not least, I also wish to express my deepest appreciation to my beloved wife and son, my mother, my father and mother in law for all over their moral support, understanding, endless love, patience and never ending encouragement and support in all ways during my study.

I certify that a Thesis Examination Committee has met on 28 June 2011 to conduct the final examination of Suliadi on his thesis entitled "**GEE-Smoothing Spline for Semiparametric Estimation of Longitudinal Categorical Data**" in accordance with the Universities and University College Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Malik bin Hj Abu Hassan, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Mohd Rizam bin Abu Bakar, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Abu Hassan Shaari Md Nor, PhD
Professor
Faculty of Economics and Management
Universiti Kebangsaan Malaysia
(External Examiner)

M. Ataharul Islam, PhD
Professor
School of Mathematical Sciences
Universiti Sains Malaysia
(External Examiner)

NORITAH OMAR, PhD
Associate Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 23 August 2011

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of **Doctor of Philosophy**. The members of Supervisory Committee were as follows:

Noor Akma Ibrahim, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Isa Daud, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Isthrinayagy S. Krishnarajah, PhD
Lecturer
Faculty of Science
Universiti Putra Malaysia
(Member)

HASANAH MOHD. GHAZALI, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia
Date:

DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

_____

SULIADI

Date: 28 June 2011

TABLE OF CONTENTS