UNIVERSITI PUTRA MALAYSIA

LIKELIHOOD INFERENCE IN PARALLEL SYSTEMS REGRESSION MODELS WITH CENSORED DATA

AYMAN. S. M. BAKLIZI

LIKELIHOOD INFERENCE IN PARALLEL SYSTEMS REGRESSION MODELS WITH CENSORED DATA

BY

AYMAN. S. M. BAKLIZI

Dissertation Submitted in Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Faculty of Science and Environmental Studies, Universiti Putra Malaysia

July 1997

TO MY PARENTS

ACKNOWLEDGEMENTS

Firstly, praise be to God, for giving me the strength and patience to complete this work. I would like to thank my supervisor Assoc. Prof. Dr. Isa Daud for his excellent supervision, invaluable guidance, helpful discussions and continuous encouragement, without which this work could not be finished. My thanks also to the members of my supervisory committee, Assoc. Prof. Dr. Mat Yusoff bin Abdullah, and Dr. Noor Akma Ibrahim for their invaluable discussions, comments, and help.

My thanks to Assoc. Prof. Dr. Harun bin Budin. His help and continuos encouragement are highly appreciated.

I wish also to express my thanks to all my friends in Malaysia and Jordan. In particular I would like to thank my dear friend Mohammed Saleh for his continuous help, encouragement, and support.

Financial support was provided by the Public Services of Malaysia (JPA), which is highly appreciated and gratefully acknowledged.

TABLE OF CONTENTS

List of Tables

LIST OF FIGURES

FIGURE **Page**

Abstract of thesis submitted to the Senate of Universiti Putra Malaysia in fulfilment of the requirements for the degree of Doctor of Philosophy.

LIKELIHOOD INFERENCE IN PARALLEL SYSTEMS REGRESSION MODELS WITH CENSORED DATA

by

AYMAN S. M. BAKLIZI

1997

Chairman : Assoc. Prof. Dr. Isa Daud

Faculty : Science and Environmental Studies

The work in this thesis is concerned with the investigation of the finite sample performance of asymptotic inference procedures based on the likelihood function when applied to the regression model based on parallel systems with censored data. The study includes investigating the adequacy of these inferential procedures as well as investigating the relative performances of asymptotically equivalent likelihood-based statistics in small samples.

The maximum likelihood estimator of the parameters of this model is not available in closed form. Thus, its actual sampling distribution is intractable. A simulation study is conducted to investigate the bias, the finite sample variance, the asymptotic variance obtained from the inverse of the observed Fisher information matrix, the adequacy of this approximate asymptotic variance, and the mean squared

error of the maximum likelihood estimator of the parameters of the regression model under consideration.

Exact hypotheses testing procedures for the model are intractable. Three standard large sample statistics based on the maximum likelihood estimator were considered. They are the Wald, the Rao, and the likelihood ratio statistics. Their performances in finite samples in terms of their sizes and powers are investigated and compared. Confidence intervals based on inverting these statistics were studied. Here again their performances in terms of the attainment of the nominal error probability and symmetry of lower and upper probabilities were investigated and compared.

The convergence of the likelihood ratio statistic to its approximating chi-squared may be improved by adjusting for a Bartlett correction factor. An alternative approach often adopted is to adjust the signed square root of the likelihood ratio statistic by the mean and variance correction. The performances of the intervals obtained from these corrections are investigated and compared. Situations under which the corrections appear to improve the quality of confidence intervals based on the likelihood ratio statistic were explained and identified.

The main findings of the simulation studies concerning likelihood inference procedures for the intercept and the slope parameters of the regression model based on parallel systems in the presence of censoring are as follows

- The variance estimates obtained from the inverse of the observed Fisher information matrix appear to be highly accurate. Estimates of the slope are nearly unbiased, while estimates of the intercept tend to be slightly biased for small to moderate sample size.

- For the hypotheses testing problem, the likelihood ratio statistic appears to perform better than the Wald and the Rao statistics.

- Interval estimates for the intercept term based on the Rao and the Wald statistic are highly asymmetric and tend to be slightly anticonservative, while intervals based on the likelihood ratio statistic are in general symmetric and attain the nominal error probability. For the slope term, all intervals tend to be symmetric for moderate to large sample size.

- The likelihood ratio statistic appear to be more suitable for one sided interval estimation and one sided hypotheses testing.

- For small sample size and high censorship level, confidence intervals based on the mean and variance correction to the signed square root of the likelihood ratio statistic appear to give accurate results; thus improving the performance of the likelihood ratio statistic.

Abstrak dissertasi yang dikemukakan kepada Senat Universiti Putra Malaysia bagi memenuhi keperluan untuk Ijazah Doktor Falsafah.

INFERENS KEBOLEHJADIAN DALAM MODEL REGRESI BERSISTEM SELARI DENGAN DATA TERTAPIS

oleh

AYMAN S. M. BAKLIZI

1997

Pengerusi: Prof. Madya Dr. Isa Daud

Fakulti : Pengajian Sains dan Alam Sekitar

Kajian di dalam tesis ini merupakan penyelidikan ke atas perlaksanaan sampel terhingga menggunakan prosedur inferens asimptot yang berdasarkan fungsi kebolehjadian apabila ia dilaksanakan ke atas model regresi yang berlandaskan sistem selari dengan data tertapis. Penyelidikan merangkumi kajian ke atas sifat kecukupan prosedur inferens dan sekaligus mengkaji perlakuan relatif bagi statistik-statistik kebolehjadian yang setara menggunakan sampel yang kecil.

Penganggar kebolehjadian maksimum bagi parameter model ini tidak boleh diperolehi secara tertutup. Oleh itu taburan pensampelan yang sebenarnya tidak boleh disingkap kembali. Kajian simulasi diperlukan untuk menjalankan dan menyelidiki kepincangan, varians bagi sampel terhingga, varians berasimptot yang dibina dari matriks informasi Fisher tercerap, kecukupan bagi hampiran varians berasimptot dan min ralat kuasa dua bagi penganggar parameter kebolehjadian maksimum dalam model regresi yang dipertimbangkan.

Prosedur ujian hipotesis yang tepat bagi model ini juga tidak boleh disingkap kembali. Lantas tiga statistik yang piawai bagi sampel besar yang berdasarkan penganggar kebolehjadian maksimum dipertimbangkan. Statistik-statistik yang berkenaan ialah statistik Wald, Rao dan nisbah kebolehjadian. Perlakuan statistik ini dalam sampel terhingga dibandingkan dari aspek saiz dan kuasa ujian. Selang keyakinan berdasarkan pendekatan songsangan statistik ini dikaji selidik. Juga Perlakuan statistik-statistik ini diselidiki dan dibanding dari segi pencapaian ralat kebarangkalian yang nominal dan kebarangkalian sebelah bawah dan atas yang simetri.

Penumpuan statistik nisbah kebolehjadian kepada hampiran khi-kuasa dua boleh diperbaiki dengan melakukan penyuaian kepada faktor pembetulan Bartlett. Satu pendekatan alternatif yang lazim diambil pakai ialah sesuaikan kuasa dua statistik nisbah kebolehjadian terhadap min dan pembetulan kepada varians. Perlakuan selang-selang yang dibina dari pembetulan ini diselidik dan dibanding. Keadaan dimana pembetulan menunjukkan peningkatan kualiti selang keyakinan berlandaskan statistik nisbah kebolehjadian dikenal pasti dan dikupas selanjutnya.

Penemuan utama yang diperolehi dari kajian simulasi berhubung dengan prosedur inferens kebolehjadian bagi parameter pintasan dan kecerunan dalam model regresi yang berdasarkan sistem selari dengan kehadiran tapisan adalah seperti berikut:-

Anggaran kepada varians yang diperolehi dari songsangan matriks informasi Fisher menunjukkan ketepatan yang sangat tinggi. Anggaran bagi parameter kecerunan hampir saksama manakala anggaran bagi parameter pintasan cenderung kepada kepincangan bagi saiz sampel yang kecil dan sederhana.

Bagi pemasalahan ujian hipotesis, statistik nisbah kebolehjadian memperlihatkan perlaksanaan yang lebih baik dari statistik Wald dan Rao.

Anggaran selang bagi parameter pintasan berdasarkan statistik Rao dan Wald memperlihatkan sifat tak simetri yang tinggi dan antikonservatif secara tidak keterlaluan. Selang-selang yang berdasarkan statistik nisbah kebolehjadian pada umumnya simetri dan mencapai ralat kebarangkalian yang nominal. Bagi parameter kecerunan pula, kesemua selang yang dibina menunjukkan kecenderungan ke arah simetri bagi saiz sampel yang sederhana dan besar.

Statistik nisbah kebolehjadian memberikan gambaran yang ianya adalah sesuai bagi anggaran selang satu hujung dan ujian hipotesis satu hujung.

Bagi saiz sampel yang kecil dengan aras tapisan yang tinggi, selang keyakinan yang dibina berdasarkan min dan pembetulan varians kepada punca kuasadua bertanda dari statistik nisbah kebolehjadian memberikan keputusan yang tepat lantas memperbaiki perlaksanaan statistik nisbah kebolehjadian ini.

CHAPTER I

INTRODUCTION

General Overview

The general purpose of statistical theory is to analyze the performance of statistical procedures, and to provide methods for the construction of optimal ones. Exact statistical theory meets these requirements in only rather special cases. In the majority of problems, either it provides a solution which is rather complicated, or an exact solution is not available at all. As an example, the sample mean is an unbiased estimator for the population mean for any distribution with finite population mean, but the exact sampling distribution of the sample mean, although known in principle, will be in an explicit usable form only for special distributions such as the normal or gamma. A second example arises in fairly complicated bayesian analyses, whenever the posterior distribution of the parameter of interest has to be obtained by high dimensional numerical integration. Other examples occur in survival and accelerated testing models, where the presence of censoring make it difficult or even impossible to work out exact solutions.

In such cases, it is necessary to rely on approximate solutions, approximate evaluation of performance, and methods for the construction of approximately optimal procedures. The so - called asymptotic theory is usually employed to handle