Small Area Estimation

via

Generalized Linear Models.

A thesis presented in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Statistics

At Massey University, Palmerston North, New Zealand.

Alasdair D. L. Noble

2003

Massey University
COLLEGE OF SCIENCES

INSTITUTE OF INFORMATION
SCIENCES & TECHNOLOGY
Private Bag 11 222
Palmerston North
New Zealand
T 64 6 356 9099
F 64 6 350 5750
www.massey.ac.nz
www-ist.massey.ac.nz

Integrated research and teaching in the fields of
- Statistics
- Computer Science
- Electronics & InfoComm Engineering

CERTIFICATE OF REGULATORY COMPLIANCE

This is to certify that the research carried out in the Doctoral Thesis entitled

Small Area Estimation via Generalized Linear Models

in the
Institute of Information Sciences and Technology
and
Statistics Research and Consulting Centre
at Massey University, New Zealand

(a) is the original work of the candidate, except as indicated by appropriate attribution in the text and/or in the acknowledgements;

(b) that the text, excluding appendices/annexes, does not exceed 100,000 words;

(c) all the ethical requirements applicable to this study have been complied with as required by Massey University, other organizations and/or committees which had a particular association with this study, and relevant legislation.

Please note Ethical Authorisation code(s) were not applicable.

Candidate's Name: Alasdair Dewar Lowe Noble

Signature:  Date: 13/12/04

Supervisor's Name: Dr Stephen Haslett

Signature:  Date 13/12/04

Massey University
COLLEGE OF SCIENCES

INSTITUTE OF INFORMATION
SCIENCES & TECHNOLOGY
Private Bag 11 222
Palmerston North
New Zealand
T 64 6 356 9099
F 64 6 350 5750
www.massey.ac.nz
www-ist.massey.ac.nz

Integrated research and teaching in the fields of
- Statistics
- Computer Science
- Electronics & InfoComm Engineering

CANDIDATE'S DECLARATION

This is to certify that the research carried out for my Doctoral thesis entitled:

"Small Area Estimation via Generalized Linear Models"

in the:

Institute of Information Sciences and Technology,
and Statistics Research and Consulting Centre
Massey University,
Palmerston North,
New Zealand

is my own work and that the thesis material has not been used in part or in whole for any other qualification.

Alasdair Dewar Lowe Noble

Signature adNoble

Date 13/12/04

Massey University
COLLEGE OF SCIENCES

INSTITUTE OF INFORMATION SCIENCES & TECHNOLOGY
Private Bag 11 222
Palmerston North
New Zealand
T 64 6 356 9099
F 64 6 350 5750
www.massey.ac.nz
www-ist.massey.ac.nz

Integrated research and teaching in the fields of
- Statistics
- Computer Science
- Electronics & InfoComm Engineering

SUPERVISOR'S DECLARATION

This is to certify that the research carried out for the Doctoral thesis entitled "Small Area Estimation via Generalized Linear Models" was done by Alasdair Noble in the Institute of Information Sciences and Technology, and the Statistics Research and Consulting Centre, Massey University, Palmerston North, New Zealand. The thesis material has not been used in part or in whole for any other qualification, and I confirm that the candidate has pursued the course of study in accordance with the requirements of the Massey University regulations.

Supervisor's Name Dr. Stephen Haslett

Signature 

Date 13/12/04

Abstract

Survey information is commonly collected to yield estimates of quantities for large geographic areas, for example, complete countries. However the estimates of those quantities at much smaller geographic areas are often of interest and the sample sizes in these areas are generally too small to give useful results. Small area estimation is used to make inference about those small areas with greater precision than the direct estimates, either by exploiting similarities between different small areas or by accessing additional information often from administrative records.

The majority of the traditional small area estimation methods are examples of a simple linear model Marker (1999) and this work begins by extending the model to a generalized linear model (GLM) Nelder and Wedderburn (1972) and then including structure preserving estimation (SPREE) in the classification. This had not been done previously.

SPREE had previously been fitted using the iterative proportional fitting algorithm Deming and Stephan (1940) which could be described as a "black box" approach. By expressing SPREE in terms of a GLM an alternative algorithm for fitting the method is developed which elucidates the underlying concepts. This new approach allows the method to be extended from the contingency table with categorical variables which the IPF could fit, to continuous variables and random effects models. An example including a continuous variable is given.

SPREE is a method which uses auxiliary information as well as survey data. In the past assumptions about appropriate auxiliary information have been made with little theoretical support. The new approach allows these assumptions to be considered and they are found to be wanting in some cases.

An example based on a national survey in New Zealand for unemployment statistics, is used extensively throughout the thesis. These data have characteristics that make analysis in the Bayesian paradigm appropriate. This paradigm has been applied and a conditional autoregressive error structure is considered.

Finally relative risk models are considered. It is shown that these could have been fitted using the IPF algorithm but the new approach allows combinations of other modeling techniques which are not available using IPF.

Acknowledgements

To fully acknowledge everyone who has had some input into this thesis would require far more space than is sensible to use. To have embarked on this journey at a mature age has meant that many people have supported me in many ways over a number of years, to all of them I am very appreciative. A few who deserve particular mention are listed below.

To all of you who have been supportive and though you may have felt I was being stupid attempting this, never voiced your thoughts; I thank you.

Firstly to Jeff Hunter and Dick Brook who were so encouraging when I first arrived at Massey University as an experienced teacher but a very naïve research student.

To the members of the then Department of Statistics and those now in the statistics group in the Institute of Information Sciences and Technology, you have been very patient, accepted my few strengths and unfailingly helped me in times of need (and there have been many). I think I have brought something to the group but know that I have taken far more.

To Doug Stirling who supervised my Masters thesis and in that way began my independent academic thinking.

To Greg Arnold who, as second supervisor, has helped me in many subtle ways. Your self deprecating manner belies a very thorough understanding of statistics and although the topic may not be very familiar to you you have always managed to relate it to areas that are familiar and in so doing bring a new light to the problem. This breadth has helped me often.

Finally on the academic side Steve Haslett. I am fairly sure when you took me on as a PhD student you had little idea of what you were letting yourself in for. I suspect the experience will make you more careful in your selection in future. I hope you have learnt a little about small area estimation through our work, I am sure that you have learnt some things about the variability in styles of learning. This work would not have been possible without you and I am certain that the benefit to me has been far greater than any benefit to you.

To Heather, Josie and (always last) Alex. I cannot thank you enough for your support through the past few years. It has had its ups and downs but hopefully the future will make it all worthwhile. At last I will be able to say to Josie and Alex "Yes I am a doctor now".

Finally I have tried to follow my Great Great Uncle Sir James Dewar's maxim:

> "Minds are like parachutes; they only function when they are open."

I am not sure that I have succeeded all of the time.

Table of Contents

Table Of Figures and Tables