

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

INTERCENSAL UPDATING OF SMALL AREA ESTIMATES

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

Marissa Cinco Isidro

2010

To Inay and Itay

Abstract

Small area estimation (SAE) involves fitting statistical models to generate statistics for areas where the sample size of the survey data is insufficient for generating precise estimates. A recent application of SAE techniques is in estimating local level poverty measures in Third World countries necessary for aid allocation and monitoring of the Millennium Development Goals (MDGs). The SAE technique commonly known as ELL method (Elbers et al., 2003) is extensively implemented by the World Bank in collaboration with national statistical agencies in most Third World countries. This technique generates estimates by fitting a linear mixed model to household level income or consumption using the survey and census data. The ELL method differs in various ways from the mainstream SAE techniques, two of which are emphasized in this thesis: (1) the ELL model does not include area level effects and (2) the model fitting technique follows a non-standard weighted generalized least squares (GLS).

Under the ELL method the survey and the census data are assumed to have been conducted at the same time period, hence generating updated estimates of poverty measures during non-census years is a problem. The method for SAE updating developed in this thesis is called the Extended Structure Preserving Estimation (ESPREE) method, an extension of the classical SAE technique called the structure preserving estimation (SPREE) method - an approach to SAE based on a categorical data analysis framework. The ESPREE method is structured within a generalized linear model (GLM) framework and uses information from the most recent survey and pseudo-census (census replicates) data to generate updated small area estimates under a superpopulation.

The World Bank in collaboration with the National Statistical Coordination Board in the Philippines has conducted an intercensal updating project using an ELL-based method requiring time invariant variables. Comparison of the estimates generated from the ELL-based and ESPREE updating method revealed substantial differences. The ESPREE method but not the ELL updating method generated unbiased estimates. An in-country validation exercise conducted in the Philippines supported the view that ESPREE based estimates, besides having theoretical advantages, also conformed better to local experts' opinion on current poverty levels.

Acknowledgements

Undertaking my PhD research requires the help of countless people and institutions, without them nothing could have been accomplished. I would like to acknowledge some of them, because to enumerate all of them is impossible.

I am very much privileged to have been accepted as a student of two of the prominent international figures and experts in small area estimation of poverty statistics in developing countries. I have learned so much from them, their valuable suggestions and constant guidance made this thesis possible. They have been very supportive in all aspects of my post graduate student life, from ensuring that my research was in the right direction, to helping me secure sufficient financial support. They are like my two most dependable sherpas in my quest to reach the highest academic mountain. To Steve Haslett and Geoff Jones, thank you so much. It has been an amazing journey!

I am so fortunate to belong to the Department of Statistics in Massey University, composed of high caliber statisticians who are very friendly and helpful, headed by the ever smiling and very approachable subject leader, Martin Hazelton. These statisticians have inspired and helped me in so many ways. To all of you, thank you very much.

Just like any research endeavor, I benefited so much from the assistance of various institutions - provision of the necessary data, academic research, and financial related support. I would like to extend my gratitude to the head and staff of the following institutions: the World Food Program (WFP), Bureau for Asia in Thailand; the National Statistics Office (NSO) and the National Statistical Coordination Board (NSCB) in the Philippines; Education New Zealand through the New Zealand Postgraduate Study Abroad Awards (NZPSAA); the Institute of Fundamental Sciences in Massey University; the International Center of Excellence for Education in Mathematics (ICE-EM) in Australia; the Cavite State University (CvSU) and the various provincial government offices in the Ilocos region of the Philippines.

In the ups and downs of my postgraduate student life, it was comforting to know that I was not alone in my struggles. My fellow Statistics postgraduate students from different parts of the world have been a source of encouragement and have made me feel “normal” most of the time! My utmost gratefulness goes to all of them especially to Ting Wang, my officemate in the writing up room.

Some of the data sets required for my research were outputs from a collaborative project of the World Bank with the NSCB. I am so fortunate to have good friends and former colleagues in NSCB whose untiring support for all my requests made my access to the needed data easier. To Dette Balamban, Glennie Amoranto, Tess Almarines, Joseph and Mildred Addawe, and Art Martinez *marami pong salamat sa inyong lahat*.

Map generation for the poverty statistics produced from my research requires some programming skills and knowledge of the necessary computer software. I managed to

save time generating the required poverty maps with the help of a former classmate and a very good friend. To Irvin Samalca, *daghang salamat Kyo!*

I would not have gone this far academically had it not been for my former lecturers and mentors who have inspired and inculcated in me an interest in the wonders of the field of Statistics. I am forever thankful to Jacqueline Guarte, Lisa Bersales, Ana Tabunda, and Erniel Barrios.

Doing research can sometimes be tiring, some rest and recreation are therefore necessary. I am so grateful to have found new friends who have helped kept my sanity and made my leisure time meaningful, full of fun and laughter. They have also opened up opportunities for me to experience different cultures, learn new language and most of all enjoy and feast on delicious food! I am very grateful to Jojo Roldan and family, Olive Pimentel and family, Cheryl Fernandez, Mimi Dogimab, Andree Wallace, May Nawanuparatsakul, Poy Theerasin, Emily Kawabata, Lala Komalawati, Leela Awaludin, Phine and Bong Flores, Edith and Roy Meeking.

Long time friends who have always been there through thick and thin. Friends who have helped me in various ways while I was doing my research. To Venus Bermudo, Badet Montana, Beng Umali, Adel Rivera, Miriam Du, Aris Magallanes, Marian Baclayon and Sanae Tacata, thank you very much.

Living *Down Under* has been bearable because I never felt alone. Up in the *Great White North* I have my best friend, who has constantly encouraged me in so many times that I was on the verge of giving up. To my husband, Phelan, saying *do tze* is not sufficient to express my gratefulness for everything you have done for me. I am also indebted to both our families for their untiring love, prayers and moral support in the past three years.

Above all, I would like to acknowledge the *Greatest Statistician* of all time. The only one who knows the true value of all the parameters on earth. Thank you so much for the gift of wisdom, good health and strength.

Palmerston North, New Zealand
31 August 2010

Marissa Isidro

Table of Contents

Abstract	iii
Acknowledgements	iv
List of Tables	xi
List of Figures	xvii
Introduction	1
Chapter 1 Small Area Estimation	8
1.1 Introduction	8
1.2 Framework for Small Area Models	9
1.3 The Linear Mixed Model	11
1.4 Small Area Estimation Techniques	12
1.4.1 Framework for SAE method	13
1.4.2 Empirical Best Linear Unbiased Prediction (EBLUP)	15
1.4.3 Empirical Bayes	17
1.4.4 Hierarchical Bayes	18
1.5 Small Area Estimation Techniques Using Survey Weights	21
1.5.1 Pseudo-Estimated Best Linear Unbiased Prediction	22
1.5.2 Iterative Weighted Estimating Equation	23
1.5.3 Pseudo-HB method	24
1.6 Summary	25
Chapter 2 Small Area Estimation of Poverty Measures	27
2.1 Introduction	27
2.2 Measures of Poverty	27
2.3 SAE for Poverty Measures in Developing Countries	30
2.3.1 The ELL Model	32
2.3.2 ELL Model Fitting Technique	34
2.3.3 Generation of Small Area Estimates of Poverty Measures	38

2.4	ELL and Other Parameter Estimation Methods	39
2.4.1	The Pseudo-EBLUP and IWEE Method	40
2.4.2	The GSR Method	41
2.4.3	Comparison of the Parameter Estimation Methods	42
2.4.4	Application to Real Data	44
2.4.5	Conclusion and Recommendations	53
2.5	Summary	55
Chapter 3	Intercensal Updating of Small Area Estimates of Poverty Measures	57
3.1	Introduction	57
3.2	Traditional Updating Methods	57
3.2.1	Demographic Methods	57
3.2.2	Indirect Procedures	58
3.3	Updating Techniques for Small Area Estimates of Poverty Measures .	60
3.3.1	Panel Data Approach in Updating Small Area Estimates . . .	61
3.3.2	Cross-sectional Surveys for Updating Small Area Estimates . .	63
3.4	Summary	65
Chapter 4	The Structure Preserving Estimation Method	66
4.1	Introduction	66
4.2	The SPREE method and IPF	67
4.2.1	The SPREE method	67
4.2.2	Iterative Proportional Fitting Algorithm	72
4.3	Loglinear Models and the SPREE method	73
4.3.1	Loglinear Models for SPREE	73
4.3.2	Model fitting for Loglinear Models	75
4.3.3	Model fitting for SPREE	77
4.4	The Logistic Regression Model for SPREE	78
4.5	Modified SPREE	80
4.6	The Generalized Linear Structural Models	81
4.7	Summary	83

Chapter 5	Extended SPREE for Updating Small Area Estimates	85
5.1	Introduction	85
5.2	SPREE as an Updating Method	85
5.3	The Extended SPREE Updating Model	87
5.4	ESPREE Modelling Procedure	90
5.5	Fitting ESPREE Models with Zero Frequencies	91
5.6	Comparison of ESPREE with Selected Methods	93
5.7	Summary	95
Chapter 6	Variance Estimation	96
6.1	Introduction	96
6.2	Variance Estimation for SPREE	97
6.3	Variance Estimation Methods for Intercensal Small Area Estimates	98
6.3.1	Linearization Method	99
6.3.2	Replication Methods	102
6.4	Summary	109
Chapter 7	Application to the Philippine Data	111
7.1	Introduction	111
7.2	The Data	111
7.2.1	Survey Data	112
7.2.2	Census Data	114
7.3	Model Formulation	115
7.3.1	The Auxiliary Variables	115
7.3.2	The Survey Margins	118
7.3.3	The Models	118
7.4	Variance Estimation	120
7.5	Illustration of ESPREE Modelling Procedure	121
7.6	Intercensal Small Area Estimates of Poverty Incidence	122
7.6.1	Municipal Level Estimates	123
7.6.2	Provincial Level Estimates	123
7.6.3	Regional Level Estimates	126
7.7	Discussion	128

Chapter 8	Validation Study	131
8.1	Introduction	131
8.2	Validation Exercise Design	132
8.2.1	Mechanics of the Validation Exercise	132
8.2.2	Areas Covered	133
8.2.3	Validation Exercise Participants	135
8.3	Validation Exercises Results	135
8.3.1	Province of Ilocos Norte	137
8.3.2	Province of Ilocos Sur	138
8.3.3	Province of La Union	142
8.3.4	Province of Pangasinan	143
8.4	Discussion and Recommendation	145
Chapter 9	Concluding Remarks	149
9.1	Introduction	149
9.2	Summary of Results and Conclusions	149
9.3	Recommendations and Future Directions	153
Bibliography		156
Appendices		163
Appendix A	Do Files in Stata for Different Survey Fitting Methods	164
Appendix B	Regional and Provincial Level Models Fitted using ELL, PEB, IWEE and GSR	204
Appendix C	Variance-covariance Structure of SPREE Estimates	231
Appendix D	Crosstabulation of Poverty Status with the Auxiliary Variables	234
Appendix E	Correlation Matrix for the Margins	241
Appendix F	Do Files in Stata for the ESPREE Updating Method	244

Appendix G Sample Validation Form	263
---	-----

List of Tables

Table 2.1	National level estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.1576633) **Based from the ELL method	48
Table 2.2	Regional level estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.18930) **Based from the ELL results	49
Table 2.3	Provincial level estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.23749) **Based from the ELL method	49
Table 2.4	Kruskal-Wallis Test for Estimated Variances at the Municipal Level (N=1243)	54
Table 2.5	Kruskal-Wallis Test for Estimated Variances at the Provincial Level (N=83)	54
Table 2.6	Kruskal-Wallis Test for Estimated Variances at the Regional Level(N=16)	54
Table 7.1	Structure of the 2000 and 2003 survey data	114
Table 7.2	Structure of 10 percent long form census	115
Table 7.3	List of auxiliary variables considered for ESPREE modelling . .	116
Table 7.4	Regression coefficient for 2000 and 2003 model for log per capita income	118
Table 7.5	Survey margins of the auxiliary variables	119
Table 7.6	Some of the models fitted with the corresponding AIC	120
Table 7.7	Summary of municipal level estimates via ESPREE and ELL .	123
Table 7.8	Summary of provincial level estimates via ESPREE and ELL .	124
Table 7.9	Comparison of regional level poverty incidence estimates	128
Table 8.1	Rank correlation between participants assessment and the small area estimates (ESPREE and ELL)	136
Table 8.2	Validation result for the province of Ilocos Norte	139

Table 8.3	Validation result for the province of Ilocos Sur	141
Table 8.4	Validation result for the province of La Union	144
Table 8.5	Validation result for the province of Pangasinan	146
Table B.1	Region 1 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.1893)	205
Table B.2	Region 2 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.1692)	206
Table B.3	Region 3 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.1478)	207
Table B.4	Region 4 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1528)	208
Table B.5	Region 5 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1578)	209
Table B.6	Region 6 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1659)	210
Table B.7	Region 7 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1899)	211
Table B.8	Region 8 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1757)	212
Table B.9	Region 9 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1727)	213
Table B.10	Region 10 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1815)	214
Table B.11	Region 11 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1906)	215

Table B.12	Region 12 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.2166)	216
Table B.13	Region 13 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1676)	217
Table B.14	Region 14 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1664)	218
Table B.15	Region 15 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1071)	219
Table B.16	Region 16 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1633)	220
Table B.17	Province 1 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.2375)	221
Table B.18	Province 2 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.1926)	221
Table B.19	Province 3 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.1565)	222
Table B.20	Province 4 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.1970)	222
Table B.21	Province 5 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.1960)	223
Table B.22	Province 6 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.1837)	223
Table B.23	Province 7 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.1130)	224

Table B.24	Province 8 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.2440)	224
Table B.25	Province 9 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.1927)	225
Table B.26	Province 10 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.1775)	225
Table B.27	Province 11 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.1677)	226
Table B.28	Province 12 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.2565)	226
Table B.29	Province 13 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1880)	227
Table B.30	Province 14 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.1427)	227
Table B.31	Province 15 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1874)	228
Table B.32	Province 16 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.2439)	228
Table B.33	Province 17 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.1898)	229
Table B.34	Province 18 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean=0.2034)	229
Table B.35	Province 19 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.2131)	230

Table B.36 Province 20 estimates of regression parameters with the standard errors and the variance components for the four techniques. *Different value for each household(mean= 0.2797)	230
Table C.1 Variance estimate of survey margins using Linearization and Half Jackknife methods	233
Table D.1 Poverty status by marital status cross-tabulation	234
Table D.2 Poverty status by strong wall material cross-tabulation	234
Table D.3 Poverty status by light wall material cross-tabulation	234
Table D.4 Poverty status by salvaged wall material cross-tabulation	235
Table D.5 Poverty status by other wall material cross-tabulation	235
Table D.6 Poverty status by strong roof material cross-tabulation	235
Table D.7 Poverty status by light roof material cross-tabulation	236
Table D.8 Poverty status by salvaged roof material cross-tabulation	236
Table D.9 Poverty status by other roof material cross-tabulation	236
Table D.10 Poverty status by male headed household cross-tabulation	237
Table D.11 Poverty status by household employs domestic helper cross-tabulation	237
Table D.12 Poverty status by living in a single type of house cross-tabulation	237
Table D.13 Poverty status by living in a duplex type of house cross-tabulation	238
Table D.14 Poverty status by living in a multiple(apartment/condominium) type of house cross-tabulation	238
Table D.15 Poverty status living in a Commercial/Industrial/Agricultural Building/House type of house cross-tabulation	238
Table D.16 Poverty status by living in other type of house (cave, boat, etc.) cross-tabulation	239
Table D.17 Poverty status by urbanity cross-tabulation	239
Table D.18 Poverty status by having a family member with college education cross-tabulation	240
Table D.19 Poverty status by having a family member (10 years old and over) with no education cross-tabulation	240
Table D.20 Poverty status by having a family member (10 years old and over) with only high school education cross-tabulation	240

Table E.1	Correlation matrix for the margins	242
Table E.2	Correlation matrix for the margins - continuation from previous page	243

List of Figures

Figure 4.1	An example of data structure for SPREE method	69
Figure 4.2	An example of data structure for SPREE method	70
Figure 4.3	An example of the allocation structure for SPREE method . .	71
Figure 7.1	Quantile-Quantile plot of ESPREE and ELL updated estimates	124
Figure 7.2	2003 Municipal Level poverty incidence estimates	125
Figure 7.3	Quantile-Quantile plot of ESPREE and FIES provincial level estimates	126
Figure 7.4	Quantile-Quantile plot of ELL and FIES provincial level estimates	127
Figure 7.5	Quantile-Quantile plot of FIES and ESPREE regional level es- timates	128
Figure 7.6	Quantile-Quantile plot of FIES and ELL regional level estimates	129
Figure 8.1	Ilocos region with ESPREE municipal level poverty incidence estimates	134
Figure G.1	Validation Form Part 1	264
Figure G.2	Validation Form Part 2	265
Figure G.3	Validation Form Part 3	266
Figure G.4	Validation Form Part 4	267