

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.

Classification Trees for Poverty Mapping

A thesis presented in partial
fulfilment of the requirements
for the degree of
Master
in Applied Statistics

at Massey University,
Palmerston North, New Zealand.

Tian Mao
2010

Abstract

Measuring differences in poverty levels within a country is important for aid allocation. Small area estimates of poverty incidence can be found by combining census and survey data. The usual method uses multiple regression, but an intuitive alternative is to build a classification tree for classifying households as poor or non-poor. This research presents some preliminary results using this method, and compares them to the traditional regression method.

Acknowledgements

To fully acknowledge everyone who has helped me throughout the year. Many people have supported me in many ways over this year, to all of them I am very appreciative. A few who deserve particular mention are listed below.

Firstly to my supervisors Geoff Jones, Siva Ganesh and Stephen Haslett. They were so encouraging and been ongoing support, guidance over the period of writing the thesis as well as ensured my Master year has been successful, educational and enjoyable. Throughout the year they have been more than a supervisor to me, they have showed me the way of having academic and independent thinking. From them I learnt to be an organized person for knowing every step of your way. This has helped me often.

To the members of the Department of Statistics and those now in the statistics group in the Institute of Fundamental Sciences. Thank you for been patient and help me in times of need.

Finally, I dedicate my work to my parents. I cannot thank them enough for their support through the years I am studying in New Zealand. It has had its ups and downs, hopefully will make it all worthwhile in the future.

Contents

Abstract	1
Acknowledgements	3
1 Introduction	6
1.1 Poverty	6
1.2 Survey Design	7
1.3 Small Area Estimation with ELL method	9
1.4 Contribution	11
2 Classification	12
2.1 Choosing a split	12
2.2 Optimizing the Tree	16
2.3 Results	18
3 Classification Trees with Weighted Data	25
3.1 Weighting by Duplication	25
3.1.1 Duplication with household Weights (Gini and Entropy)	26
3.1.2 Duplication with per-person Weights (Gini and Entropy)	29
3.1.3 Prediction Comparison between Duplication method and SAE results	31
3.2 Weights Argument in RPART	34
3.3 Setting up priors and obtaining ROC curve	35
3.4 Comparison between Confusion Matrices	36
4 Estimating the Standard Errors	42
4.1 Calculation Method without consider the Municipal Variances	42
4.1.1 Standard Error Calculation Method One	42
4.2 Calculation Methods count in Municipal Variances	48
4.2.1 Standard Error Calculation Method Two	48

CONTENTS	5
4.2.2 Standard Error Calculation Method Three	50
5 Conclusions and Further work	58
Appendix	61
List of X Variables for the Philippines data	62
Full Trees Built with Various Data Base	63
Bibliography	73