FUZZY PARAMETRIC AND FUZZY SEMI-PARAMETRIC SAMPLE SELECTION MODELS

MUHAMAD SAFIH BIN LOLA

UNIVERSITI SAINS MALAYSIA
2011

FUZZY PARAMETRIC AND FUZZY SEMI-PARAMETRIC SAMPLE SELECTION MODELS

MUHAMAD SAFIIH BIN LOLA

UNIVERSITI SAINS MALAYSIA
2011

FUZZY PARAMETRIC AND FUZZY SEMI-PARAMETRIC SAMPLE SELECTION MODELS

by

MUHAMAD SAFIIH BIN LOLA

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

JANUARY 2011

Acknowledgements

In the Name of ALLAH s.w.t, the Most Beneficent, the Most Merciful. All praise is due only to ALLAH s.w.t., the Lord of the Universe. Ultimately, only ALLAH s.w.t. gives us the strength and courage to proceed with life in its entirety. His work is truly splendid and wholesome. His knowledge is truly complete with due perfection. I am highly indebted to many people for their assistance in the preparation of this thesis. I wish to express my sincere gratitude to my supervisor, Assoc. Prof. Dr. Anton Abdul Basah Kamil for his excellent supervision, invaluable guidance, and helpful discussions and continuous encouragement. His enthusiasm and thoughtful considerations, and his valuable suggestions throughout the preparation of this thesis are greatly appreciated. I would also like to extend my sincere gratitude and appreciation to my co-supervisors, Prof. Dr. Abu Osman Bin Md Tap and Dr. Zainuddin Bin Arshad, for their valuable suggestions and guidance.

I would also like to acknowledge with much appreciation the crucial role of the editors and referees or reviewers for all of my publications. Their valuable comments and suggestions throughout the preparation of various publications have helped me to improve my work and are highly appreciated. My special thanks to University Malaysia Terengganu (UMT), the Jabatan Perkhidmatan Awam (JPA) and the Ministry of Higher Education for the financial support through out the course of my study. I would also extend my thanks to Lembaga Penduduk dan Pembangunan Keluarga Negara (LPPKN) for data provided. I would also like to express my appreciation to my family: my dearest wife, who always provides the encouragement and support that I truly need; my loving children, who have always been the source of inspiration and pleasure; my parents, brothers and sisters for their continuous prayers for my success. Last but not least, my sincere thanks to all my colleagues and friends who kindly provided valuable and helpful comments in the preparation of this thesis, specifically Dr. Abd. Fatah, Dr. Gobithasan, Dr. Tengku Sifzizul and to those involved directly or indirectly in the preparation of this thesis, whom I have not mentioned.

TABLES OF CONTENTS

CHAPTER 3: FUZZY SETS IN A FUZZY MODELING

CHAPTER 4 : FUZZY PARAMETRIC SAMPLE SELECTION MODEL (FPSSM) AND FUZZY SEMI-PARAMETRIC SAMPLE SELECTION MODEL (FSPSSM)

CHAPTER 5: SIMULATION STUDY ON FUZZY PARAMETRIC AND FUZZY SEMI-PARAMETRIC SAMPLE SELECTION MODELS

APPENDICES

LIST OF PUBLICATIONS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

PSSM	-	Parametric sample selection model
SPSSM	-	Semi-parametric sample selection model
FPSSM	-	Fuzzy parametric sample selection model
FSPSSM	-	Fuzzy semi-parametric sample selection model
TFN	-	Triangular fuzzy number
DWADE	-	Density weighted average derivative estimator
MPFS	-	Malaysian Population and Family Survey
OLS	-	Ordinary least square
MLE	-	Maximum likelihood estimator
SIM	-	Single index model
FLR	-	Fuzzy linear regression
S.D	-	Standard deviation
MSE	-	Mean square error
RMSE	-	Root mean square error
LHS	-	Left hand side
RHS	-	Right hand side
$A\tilde{G}E$	-	The married women's age (in years) divided by 10
$A\tilde{G}E2$	-	The age squared divided by 100
EDU	-	The educational levels, measured in years of schooling
$CHILD$	-	The number of children younger than 18 years old and living in the family
$H\tilde{W}S$	-	The log of the husband's monthly wage (measured in Ringgit Malaysia)
$\tilde{Z}_i^*$	-	The log of women's hourly wage rate (measured in Ringgit Malaysia)
$P\tilde{E}XP$	-	The potential work experience, defined as (age) – 6 – (total years of schooling)
$PE\tilde{X}P2$	-	The potential work experience squared
$PEX\tilde{P}CHD$	-	The potential work experience times the number of children
$PEX\tilde{P}CHD2$	-	The potential work experience squared, times the number of children

LIST OF SYMBOLS

$\{v_1,v_2,v_3,\ldots,v_n\}$	-	Set of elements $v_1,v_2,v_3,\ldots,v_n$
$(v_1,v_2,v_3,\ldots,v_n)$	-	n-tuple
$\|\|\cdots\|\|$	-	Euclidean norm
min $(v_1,v_2,v_3,\ldots,v_n)$	-	Minimum of $v_1,v_2,v_3,\ldots,v_n$
max $(v_1,v_2,v_3,\ldots,v_n)$	-	Maximum of $v_1,v_2,v_3,\ldots,v_n$
$X_1 < X_2 < \ldots < X_n$	-	Order statistics
[a, b]	-	Closed interval of real number between a and b
[a, b)	-	Interval of real number closed in a and open in b
(a, b)	-	Open interval of real numbers between a and b
(a, b]	-	Interval of real number open in a and closed in b
$\Re$	-	Set of real number
$\Re^n$	-	Set of n-tuple or real numbers
$\aleph$	-	Set of positive integers (real numbers)
U or X	-	Universal set
A, B, C,...	-	Arbitrary sets (crisp or fuzzy)
$A = B$	-	Set equality
$A \neq B$	-	Set inequality
$A \subset B$	-	Proper set inclusion
$A \subseteq B$	-	Set inclusion
$A \cap B$	-	Set intersection
$A \cup B$	-	Set union
A'	-	Set complement
$A \times B$	-	Cartesian product of sets A and B
$\emptyset$	-	Empty set
$f: X \rightarrow Y$	-	Function of f from X into Y
$f^{-1}: X \rightarrow Y$	-	Inverse function of f from X to Y
$\Rightarrow$	-	Implies
$\Leftrightarrow$	-	If and only if
$\in$	-	Element of

MODEL PEMILIHAN SAMPEL BERPARAMETRIK KABUR DAN SEPARA BERPARAMETRIK KABUR

ABSTRAK

Sejak 30 dekad yang lalu, model regresi mendapat perhatian yang munasabah dan menunjukkan kejayaan jika diaplikasikan dengan model lain. Satu daripada model yang paling berjaya ialah model pilihan sampel atau model terpilih. Model ini merupakan gabungan daripada model probit dan model regresi. Kajian awal tentang model ini tertumpu pada pendekatan parametrik. Pendekatan standard bagi anggaran model pilihan sampel menunjukkan keputusan yang tidak konsisten sekiranya andaian taburan ralat dilakukan. Oleh itu, kemajuan yang dianggap penting dalam beberapa dekad kebelakangan ini dalam membangunkan suatu pendekatan alternatif untuk mengatasi masalah ini adalah melalui penggunaan kaedah parametrik separa. Walau bagaimanapun, ketidaktentuan dan ketaksaan wujud di dalam model ini, terutamanya perkaitan di antara pemboleh ubah endogen dan eksogen. Oleh itu, ia akan menganggu kemampuan dan keberkesanan model dalam memberikan nilai teranggar yang boleh menjelaskan situasi sebenar sesuatu fenomena. Persoalan dan masalah ini perlu diteroka, dan merupakan tujuan utama kajian ini. Suatu rangka kerja baru bagi anggaran model pilihan sampel yang menggunakan konsep pemodelan kabur diperkenalkan. Melalui pendekatan ini, suatu hibrid konsep kabur yang fleksibel dengan model-model pilihan sampel berparametrik dan berparametrik separa, yang dikenali sebagai model pilihan sampel berparametrik kabur (*fuzzy parametric sample selection model*, FPSSM) dan model pilihan sampel berparametrik separa kabur (*fuzzy semi-parametric sample selection model*, FSPSSM). Elemen kekaburan dan ketidaktentuan dalam binaan model adalah suatu

cara bagi meningkatkan perolehan maklumat untuk menghasilkan suatu model yang lebih tepat. Hal ini mendorong pembangunan teorem capahan dalam bentuk nombor kabur segi tiga yang digunakan dalam model. Di samping itu, teorem yang berkaitan dengan kecapahan dan kenormalan FPSSM dan FSPSSM juga dibangunkan. Bukti teorem dibentangkan. Dengan itu, suatu algoritma yang menggunakan konsep pemodelan kabur dibangunkan. Ketekalan merupakan suatu petunjuk bagi keberkesanan model yang dibangunkan dan dibuktikan menggunakan simulasi Monte Carlo. Untuk memudahkan pelaksanaan algoritma ini suatu pengaturcaraan yang menggunakan bahasa R dibangunkan. Ketekalan dan keefisienan model yang dicadangkan, diberi pertimbangan sewajarnya dalam kajian ini. Dalam usaha tersebut, simulasi Monte Carlo digunakan. Jadi, ralat FPSSM dan FSPSSM dianggap mematuhi taburan normal dan taburan khi kuasa dua. Di samping itu, keadaan di bawah FSPSSM serta parameter lebar jalur juga dipertimbangkan. Keputusam simulasi menunjukkan bahawa FPSSM dan FSPSSM adalah tekal dan efisien apabila taburan mereka adalah normal dan mematuhi parameter lebar jalur. Sebaliknya, FPSSM tidak stabil bagi taburan khi kuasa dua. Keputusan menunjukkan bahawa penganggar bagi FPSSM dan FPSSM adalah konsisten dan meningkatkan nilai "α-cuts", nilai penganggar yang mendekati nilai penganggar pada model asal. Selanjutnya, keberkesanan penganggar bagi model ini dikaji untuk menentukan faktor yang mempengaruhi keputusan wanita yang telah berkahwin dalam menyertai pasaran buruh. Bagi tujuan tersebut, data daripada Kajian Keluarga dan Penduduk Malaysia 1994 digunakan. Dapatan kajian menunjukkan bahawa keputusan mereka secara signifikannya dipengaruhi EDU. *EDU*, $P\tilde{E}XP$ dan $PE\tilde{X}\tilde{P}CHD$ adalah faktor penting yang mempengaruhi upah yang diterima mereka.

FUZZY PARAMETRIC AND FUZZY SEMI-PARAMETRIC SAMPLE SELECTION MODELS

ABSTRACT

In the past thirty decade, regression model has received considerable attention and has been shown to be successful if applied together with other models. One of the most successful models is the sample selection model or the selectivity model. This model is a combination of the probit and regression models. The earlier studies on this model focused on the parametric approach. The standard approach of estimating sample selection model shows inconsistent results if the distributional assumptions of the errors terms are made. Hence, an important progress within the last decade in the development of an alternative approach to overcome this problem is through the use of semi-parametric method. However, the uncertainties and ambiguities exist in the models, particularly the relationship between the endogenous and exogenous variables. Therefore, it will disrupt the ability and effectiveness of the model proceeded to give the estimated value that can explain the actual situation of a phenomenon. These are questions and problems that have yet to be explored and the main pillar of this study. A new framework for estimation of sample selection model using the concept of fuzzy modeling is introduced. Through this approach, a flexible fuzzy concept hybrid with the parametric and the semi-parametric sample selection models known as fuzzy parametric sample selection model (FPSSM) and fuzzy semi-parametric sample selection model (FSPSSM). The elements of vagueness and uncertainty in the models are represented in the model construction, as a way of increasing the available information to produce a more accurate model. This led to the development of the convergence theorem presented in the form of triangular fuzzy numbers to be used in the model. Besides that, theorems related to the

convexity and normality of the FPSSM and FSPSSM are developed. Proofs of the theorems are presented. With that, an algorithm using the concept of fuzzy modelling is developed. Consistency is an indicator of effectiveness of the developed models and is proved using Monte Carlo simulation. To facilitate the implementation of these algorithms, the programming using the *R* language is developed. The consistency and efficiency of the proposed model are considered under this study. In order to achieve that condition, a Monte Carlo simulation is used. Hence, the errors terms of FPSSM and FSPSSM are assumed to follow the normal and the chi-square distribution. In addition of those conditions under FSPSSM, the bandwidth parameters are considered. Simulation results show that FPSSM and FSPSSM are consistent and efficient when their distributional are normal and follow the bandwidth parameter. Instead, the FPSSM by chi-square distribution is found inconsistent. Results show that the estimators for FPSSM and FPSSM is consistent and increasing the α - cuts values, the values of estimator will approach to the values of estimator on the original model. Further, the effectiveness of the estimators for these models are investigated to determine the factors that cause married women in the labour market to decide when to participate or not and wage earning of married women in the labour market. For that purpose, the Malaysian Population and Family Survey 1994 is used. The findings show that married women's decision to participate in the labour market is significantly influenced by EDU. While *EDU*, $P\tilde{E}XP$ and $PE\tilde{X}\tilde{P}CHD$ are the most important factors affecting the wage earning by married women in the labour market.