Fuzzy Semi-parametric Sample Selection Model for Participation of Married Women

L. MUHAMAD SAFIIH [1]*, A.A.BASAH KAMIL[2], M. T. ABU OSMAN [3]
[1, 3]Mathematics Department
Faculty of Science and Technology, University Malaysia Terengganu
21030 Kuala Terengganu, Terengganu, MALAYSIA.

[2]School of Distance Learning
Universiti Sains Malaysia
11800 USM Penang, MALAYSIA.

Abstract: The sample selection model is studied in the context of semi-parametric methods. The issue of uncertainty and ambiguity are still major problems and the modelling of a semi-parametric sample selection model as well as its parametric is complicated. The best approach of accounting for uncertainty and ambiguity is to take advantage of the tools provided by the theory of fuzzy sets. The semi-parametric of a sample selection model is an econometric model that has found an interesting application in empirical studies. In this paper, the married women participants in the Malaysia labour force are studied. It comprises the analysis of a) participation equation in the wage sector and b) the wage equation in the wage sector. The data set used for this study is from the Malaysian population and family survey 1994 (MPFS-1994).

Key-words:- uncertainty, semi-parametric sample selection model, participant equation, wage equation, fuzzy number.

1 Introduction

The study of semi-parametric sample selection models has received considerable attention from statisticians as well as econometricians in the late 20th century (see Schafgans, 1996). The termed "semi-parametric," used as a hybrid model for the selection models, which do not involve parametric forms on error distributions; hence, only the regression function part of the model of interest is used. Consideration is based on two perspectives, firstly; no restriction of estimation of the parameters of interest for the distribution function of the error terms, secondly; restricting the functional form of heteroskedasticity to lie in a finite-dimensional parametric family (Schafgans,1996). Gallant and Nychka (1987) studied these methods in the context of a semi-nonparametric maximum likelihood estimation and applied the method to nonlinear regression with sample selection model. Newey (1988) used a series approximation to the selection correction term that considered regression s-pline and power series approximations. Robinson (1988) focused on the simplest interesting setting of multiple regressions with independent observations, and described extensions to other econometric models, in particular seemingly unrelated and nonlinear regressions, simultaneous equations, distribution lags and sample selectivity models. More specifically, none of these researchers put any efforts into studies that analysed semi-parametric sample selection models in the context of fuzzy environment like fuzzy sets, fuzzy logic or fuzzy sets and systems (M.Safiih (2007)).

This paper introduces a membership function of a semi-parametric sample selection model in which historical data contains some uncertainty. The concept of fuzzy sets by Zadeh is extended from the crisp sets, that is the two-valued evaluation of 0 or 1, {0, 1}, to the infinite number of values from 0 to 1, [0, 1]. (see Terano *et.al*. 1994). This provides an ideal framework to deal with problems in which there does not exist a definite criterion for discovering what elements belongs or does not belong to a given set (Miceli, 1998). Fuzzy set are defined by a fuzzy sets in a universe of discourse U is characterised by a membership function denoted by the function μ_A maps all elements of U that take the values in the interval [0,1] that is $A : X \rightarrow [0,1]$ (Zadeh, 1965).

2 Data description

The data set used for this study is from the Malaysian population and family survey 1994 (MPFS-1994). The original MPFS-94 sample data comprises 4444 married women. Based on the sequential criteria (Mroz,1984), this analysis was limited to the completed information provided by the married women. The resulting sample data set consisted of only 1100 married women. This accounted for 39.4% who were employed and the rest were considered as non-participants amounting to 1692 (60.6%). The total data sets used in this study consisted of 2792 married women. The selection rules (Martins, 2001) were applied to create the sample criteria of choosing the participant and non-participant married women based on the MPFS-94 data set.

2.1 Variables used in the study

The first equation, which is the probability that a married women participates in the labour market, the so-called participation equation. The independent variables involved AGE (age in year divided by 10), AGE2 (age square divided by 100), EDU (years of education), CHILD (the number of children under 18 living in the family), HW (log of monthly husband's wage). For the potential experience, the calculation is given by age-edu-6 rather then actual work experience. In order to deal with these problems the solution adopted uses a method advanced by Buchinsky (1998). The second equation is called wage equation. The dependent variable used for the analysis was the log hourly wages (z). The independent variables were EDU, PEXP (potential work experience divided by 10), PEXP2 (potential experience squared divided by 100), PEXPCD (PEXP interacted with the total number of children) and PEXPCHD2 (PEXP2 interacted with the total number of children).

3 Empirical results: semi-parametric and fuzzy semi-parametric SSM

This section presents the results that applies to the most basic one i.e. the participant and wage equation of DWADE estimator (Hardle *et.al*. 1995) and Powell estimator (Powell,1987), respectively. Both estimators are consistent with $\sqrt{n}-$consistency and asymptotic normality. The discussion focuses on the participation and wage equation in terms of the estimated coefficient, the significant effect and consistency of the estimate for SPSSM, as well as FSPSSM for comparison purposes.

3.1 Participation equation in the wage sector

The participation equation using the DWEDE estimator (Hardle *et.al*. 1995) is presented in Table 1 and FSPSSM results as well for comparison purposes. The first column used DWADE estimator with values of bandwidth h= 0.2 without the constant terms, followed with fuzzy semi-parametric sample selection model (FPSSM) with $\alpha -$cuts 0.0, 0.2, 0.4, 0.6 and 0.8 respectively. At first, the estimate coefficient suggests that all variables except variable age are significant (significantly and negative estimated coefficient on AGE2 and CHILD, while positive and significant coefficient estimated for EDU and HW). However, only CHILD show a statistically significantly effect at the 5% level. This was unexpected and shown to be an important result. Although in the conventional parametric model, it appears together with EDU but in the context of SPSSM, it only estimates the CHILD effect and appears significantly to be relevant, which is more in line with economic theory.

For comparison purposes, the FSPSSM is used. The estimated coefficient gives a similar trend with the SPSSM i.e. significant for variables AGE2, EDU, CHILD and HW. The results show a significantly quite and positive coefficient estimate for EDU and HW, significantly quite but negative estimated coefficient on AGE2 and CHILD. In the FSPSSM context, the CHILD coefficient appears to be relevant being statistically significant at the 5% level. This is an interesting finding and it should be pointed out when using this approach that almost the standard errors for the parameter were much smaller when compared to those in the conventional SPSSM. This is evidence that this approach is much better in estimating coefficient and gives considerable efficiency gain compared to those in the conventional semi-parametric model. In addition, the coefficient estimated obtained from FSPSSM agrees quite closely with the coefficient estimate from conventional SPSSM. Hence, the coefficient estimated from FSPSSM is consistent, even though it involves uncertainties in its data.

3.2 The wage equation in the wage sector

The wage equation using the Powell estimator (Powell, 1987) of SPSSM is presented in Table 2 and FSPSSM results for comparison purposes. The first column used Powell estimator with values of bandwidth h= 0.2 without the constant terms. Following column is given by fuzzy semi-parametric sample selection model (FPSSM) with $\alpha -$cuts 0.0, 0.2, 0.4, 0.6 and 0.8 respectively.

At first, the estimate coefficient suggested that the all variables are significant (quite significant and negative estimated coefficient on EDU, PEXP2 and PEXPCHD, while positive and significant coefficient estimated for PEXP and PEXPCHD2). As the estimated coefficient, the results for all variables gives a statistically significantly effect at the 5% level. This was given as a significant result. The results reveal the significant differences of SPSSM against PSSM methods of correcting the sample selectivity bias. This increased the results that were obtained in SPSSM where not all variables in PSSM contribute significantly for married women involved in wage sectors.

For comparison purposes, it was then applied with the FSPSSM. The estimated coefficient was significant for all variables. The results show significant and positive coefficient estimate for PEXP and PEXPCHD2, significant but negative estimated coefficient on EDU, PEXP2 and PEXPCHD. The coefficient for all variables appears to be relevant with statistical significance at the 5% level. It should be pointed out that, the standard errors for the parameter EDU, PEXP and PEXP2 were much smaller when compared to those in the conventional SPSSM. This is evidence that this method is considerably more efficient then those in the conventional semi-parametric model. The coefficient estimated obtained from FSPSSM is also considerably quite close with the coefficient estimated from conventional SPSSM. In other words, in applying FSPSSM, the coefficient estimated is consistent even though data used involved uncertainties.

4 Conclusion

For comparison of the participant equation, the estimated coefficient and the significant factor gives a similar trend with the SPSSM. However, one of the interesting findings and the most significant result appears by applying the FPSSM i.e. the FSPSSM is a better estimate when compared to the SPSSM in terms of the standard error of the coefficient estimate. The standard errors of coefficient estimate for the FSPSSM almost give a smaller when compared to the conventional SPSSM. This is evidence that this approach is much better in estimating coefficient and has a considerable efficiency gain then those in the conventional semi-parametric model. The coefficient estimated obtained was also considerably closer with the coefficient estimated by conventional SPSSM. Hence, this gives evidence that the coefficient estimated is consistent although data used involved uncertainties. Secondly, wages equation, in terms of the coefficient estimation and significant factor, the FSPSSM is considerably closer to the standard error of SPSSM. As a whole, the FSPSSM gave a better estimate when compared to the SPSSM. In terms of consistency, it was found that the coefficient estimate for all variables of FSPSSM were not much different to the coefficient estimate of SPSSM, even though the values of the $\alpha - cuts$ increased (from 0.0 to 0.8). In the other words, by looking at the coefficient estimate and consistency, the fuzzy model (FSPSSM) is much better than the model without fuzzy (PSSSM) for the participation and wage equation.

References:

Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. Journal of Applied Econometrics. Vol. 13. p. 1-30.

Gallant, R. and D. Nychka (1987). Semi-Nonparametric Maximum Likelihood estimation, Econometrica, 55, p. 363-390.

Härdle, W. Klinke, S. and Müller, M. (1999). Xplore learning guide. Berlin. Springer-Verlag.

L. M. Safiih Lola (2007). Fuzzy Semi-parametric of a Sample Selection Models. Ph.D. dissertation. University Science of Malaysia. Penang. Malaysia.

Miceli, D. (1998). Measuring poverty using fuzzy sets, Discussion paper no.38, National centre for social and economic modeling, University of Canberra.

Martin, M.F.O. (2001). Parametric and semiparametric estimation of sample selection models: An empirical application to the female labor force in Portugal. Journal of Applied Econometrics, 16, p. 23-39.

Mroz, T.A. (1984). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. Ph.D. dissertation, Stanford University London, UK.

Newey, W. (1988). Two step series estimation of sample selection models, Department of Economic, MIT working paper no. E52 - 262D. p. 1-17.

Powell, J.L. (1987). Semiparametric estimation of bivariate latent variable models. Social Systems Research Institute. University of Wisconsin-Madison, Working paper No.8704.

Robinson, P.M. (1988). Root-N consistent semiparametric regression, Econometrica, 56, p.931-954.

Schafgans, M. (1996). Semiparametric estimation of a sample selection model: estimation of the intercept; theory and applications, Ph.D. dissertation, Yale University New Haven, USA.

Terano, T. Asai, K. Sugeno, M. (1994). Applied fuzzy systems, Cambridge. AP Professional.

Zadeh, L.A. (1965). Fuzzy Sets and systems. In: Fox, J., ed., System Theory. Brooklyn, New York. Polytechnic Press.

Table 1: ***Semi-parametric and fuzzy semi-parametric estimates for the participation equation***

Participation equation	*Coefficients*					
	DWADE	*Fuzzy Selection Model*				
		α = 0.8	*α = 0.6*	*α = 0.4*	*α = 0.2*	*α = 0.0*
AGE	-0.002048 (1.233)	-0.0015393 (1.150)	-0.0043978 (1.151)	-0.0015934 (1.234)	-0.0016184 (1.232)	-0.001642 (1.232)
AGE2	-0.00016099 (0.1754)	-0.00016584 (0.1624)	-0.00020722 (0.1627)	-0.00016629 (0.1763)	-0.00016651 (0.1765)	-0.00016673 (0.1767)
EDU	0.00034766 (0.02116)	0.00023044 (0.02015)	0.00011323 (0.02015)	0.00023044 (0.02115)	0.00023044 (0.02062)	0.00023044 (0.02062)
CHILD	-0.0039216* (0.06573)	-0.0044301* (0.06341)	-0.0048986* (0.0634)	-0.0044301* (0.06571)	-0.0044301* (0.06485)	-0.0044301* (0.06484)
HW	0.044008 (0.1632)	0.050262 (0.1402)	0.05597 (0.1396)	0.049549 (0.1485)	0.049189 (0.1437)	0.048832 (0.1432)

** 5% level of significant*

Table 2: ***Semi-parametric and fuzzy semi-parametric estimates for the wage equation***

Wage equation	*Coefficients*					
	POWELL	*Fuzzy Selection Model*				
		α = 0.8	*α = 0.6*	*α = 0.4*	*α = 0.2*	*α = 0.0*
EDU	-0.0112792 (0.005262)	-0.0109003 (0.005258)	-0.010939 (0.005258)	-0.011346 (0.005259)	-0.011385 (0.005259)	-0.0114256 (0.005258)
PEXP	0.544083* (0.1099)	0.540864* (0.1096)	0.538776* (0.1094)	0.534385* (0.1093)	0.532247* (0.1092)	0.530069* (0.109)
PEXP2	-0.160272* (0.02633)	-0.159762* (0.0263)	-0.159524* (0.0263)	-0.158781* (0.02632)	-0.158525* (0.02632)	-0.158259* (0.02632)
PEXPCHD	-0.161205* (0.02453)	-0.159863* (0.02453)	-0.159583* (0.02455)	-0.15889* (0.02459)	-0.158584* (0.02461)	-0.158262* (0.02463)
PEXPCHD2	0.046591* (0.008485)	0.0463242* (0.008485)	.0462221* (0.008493)	0.0458118* (0.008508)	.0457004* (0.008511)	0.0455835* (0.008517)

** 5% level of significant*