FUZZY MODIFIED SEMI-PARAMETRIC OF SAMPLE SELECTION MODELS

[1]Yaya Sudarya Triana, [2]Sabri bin Ahmad

[1,2]Mathematics Department
Faculty of Science and Technology, University Malaysia Terengganu
21030 Kuala Terengganu, Terengganu, Malaysia
[1]yst_2000@yahoo.com, [2]sba@umt.edu.my

Abstract

The modified semi-parametric for non-participation of sample selection models are emphasized by the use of the bandwidth parameters. This paper provides a consistent and asymptotically normal estimator for the estimated of modified semi-parametric sample selection models. Some results of Monte Carlo, to demonstrate of the limited sample performance is provided. However, the problem of uncertainty and ambiguity occurred in the proposed models particularly the relationship between endogenous and exogenous variables. This paper needs to consider the model estimation using fuzzy modelling approach, called fuzzy modified semi-parametric sample selection model (FMSPSSM), the fuzzy concept and its properties is the best alternative to overcome this problem. The triangular fuzzy number with membership function is used. In the development of the proposed models, the consistency and efficiency become an important aspect and need to study. Thus, to test the consistency and efficiency of the proposed models, the Monte Carlo simulation developed by Nawata (1994) is used. In this simulation, sample size of 100, 200, and 400 are used. The results revealed that the proposed models of FMSPSSM are consistent and efficient.

Key words - Fuzzy, Monte Carlo, Powell, Sample selection model, Semi-parametric.

1 INTRODUCTION

The semi-parametric model characterized as having a finite-dimensional parameter of interest (parametric) and a finite-dimensional nuisance parameter (non-parametric) was given by Begun et al (1983). Even though this difference is a defining characteristic of semi-parametric modelling. Definitions of "parametric", "semi-parametric" and "non-parametric" according to Powell's (1994). The same estimator can be viewed as parametric, non-parametric or semi-parametric, in many cases, depending on the assumptions of the model. The classical linear model as follows :

$$y_i = x_i^{'}\beta + u_i \tag{1}$$

The unknown coefficients β is estimated by the least squares estimator :

$$\hat{\beta} = \left[\sum_{i=1}^{N} x_i x_i^{'}\right]^{-1} \sum_{i=1}^{N} x_i y_i \tag{2}$$

2 MODIFIED SEMI-PARAMETRIC OF SAMPLE SELECTION MODELS

The modified semi-parametric sample selection model (MSPSSM) as follows :

$$
\begin{aligned}
y^*_{i_{sp}} &= g(x_i) + g(w_i\alpha) + u \\
d^*_{i_{sp}} &= w'_{i_{sp}}\alpha + v_{i_{sp}} \\
y_{i_{sp}} &= y^*_{i_{sp}} d_{i_{sp}} \\
d_{i_{sp}} &= \begin{cases} 1 & \text{if} \quad d^*_{i_{sp}} \le 0 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
\tag{3}
$$

Where $y_{i_{sp}}$ and $d_{i_{sp}}$ are dependent variables, $x_{i_{sp}}$ and $w_{i_{sp}}$ are vectors of exogenous variables, α and β are unknown parameter vectors, $u_{i_{sp}}$ and $v_{i_{sp}}$ are error terms.

The MSPSSM as described in Equation (1) consist of two steps. The first step is an estimation of the participation equation using DWADE and the second step is an estimation of structural equation using Powell (1987) estimator.

The First Step: Estimation of the Participation Equation
Powell et al. (1989) proposed Density Weighted Average Derivative Estimator (DWADE) to estimate parameter α in the first step of Equation (3). DWADE that is non-iterative and easily computed. DWADE is based on sample analogues of the product moment representation of the average derivations and is constructed using non-parametric kernel estimators of the density of the regressors. To estimate of the coefficients in index models through the estimation of the density-weighted average derivative of a general regression function.

The Second Step: Estimation of the Structural Equation
Powell (1987) has proposed that the Powell estimator is used to estimate β in the second step of Equation (3). Powell (1987) regarded as a semi-parametric selection model which combines the two-equation structure with the weak distributional assumption of the joint distribution of the error terms as follows :

$$f(u_i, v_i \mid x_i) = f(u_i, v_i \mid x'_i\beta) \tag{4}$$

where $f(\cdot)$ is unknown function of the joint density of u_i, v_i (conditional on x_i).

the estimator proposed by Powell (1987) as follows :

$$\hat{\beta}_{Powell} = \left[\binom{n}{2}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\hat{\mu}_{ij}N(x_i - x_j)(x_i - x_j)'\right]^{-1} \cdot \left[\binom{n}{2}^{-1}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\hat{\mu}_{ij}N(x_i - x_j)(y_i - y_j)\right] \tag{5}$$

where $\hat{\mu}_{ij}N = \frac{1}{h}K\left(\frac{w'_i\hat{\alpha} - w'_j\hat{\alpha}}{h}\right)$ with symmetric kernel function $K(\cdot)$, $\hat{\mu}_{ij}$ is weights and are calculated, $\hat{\beta}$ is parameter estimated and obtained by a weighted least-squares estimator, h is bandwidth, $\hat{\alpha}$ is parameter estimated.

3 DEVELOPMENT OF FUZZY MSPSSM

This fuzzy concept is applied to the MPSSM and MSPSSM. Fuzzification is a process to convert non-fuzzy variables into fuzzy variables. In the first stage, this data will be in the fuzzification of input and output data. Input data are obtained from the survey of The Malaysian population and family survey 1994 (MPFS-1994) or the actual data and also from the simulation data. The output data are in the forms of levels of self-defined levels. Here the level of those levels can be implemented in the form of indicators that have been given an explanation of each relevant level. The α-cut method with an increment value of 0.1 starting with 0 up to 0.8 is then applied to the triangular membership function. From the α - cut method, a lower and upper bound for each observation are obtained (x_i, y^*_i and w_i) which is defined as

$$\tilde{x}_i = (x_{il}, x_{im}, x_{iu}),\ \tilde{y}^*_i = (y_{il}, y_{im}, y_{iu}) \text{ and } \tilde{w}_i = (w_{il}, w_{im}, w_{iu})$$

The functions of membership are as follows :

$$\mu_{\tilde{x}_i}(x) = \begin{cases} \dfrac{(x - x_{il})}{(x_{im} - x_{il})} & \text{if } x \in [x_{il}, x_{im}] \\ 1 & \text{if } x = x_{im} \\ \dfrac{(x_{iu} - x)}{(x_{iu} - x_{im})} & \text{if } x \in [x_{im}, x_{iu}] \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$$\mu_{\tilde{y}^*_i}(y) = \begin{cases} \dfrac{(y - y_{il})}{(y_{im} - y_{il})} & \text{if } y \in [y_{il}, y_{im}] \\ 1 & \text{if } y = y_{im} \\ \dfrac{(y_{iu} - y)}{(y_{iu} - y_{im})} & \text{if } y \in [y_{im}, y_{iu}] \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

$$\mu_{\tilde{w}_i}(w) = \begin{cases} \dfrac{(w - w_{il})}{(w_{im} - w_{il})} & \text{if } w \in [w_{il}, w_{im}] \\ 1 & \text{if } w = w_{im} \\ \dfrac{(w_{iu} - w)}{(w_{iu} - w_{im})} & \text{if } w \in [w_{im}, w_{iu}] \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

To create the MPSSM and MSPSSM, then the first step is to change the membership (converting real-triangular fuzzy membership values into a crisp value). A centroid method or the centre of gravity method is used to calculate the outputs of the crisp value as the centre of the area under the curve. The values of X_{ic}, Y_{ic}, and W_{ic} be the defuzzified values of $\tilde{X}_i$, $\tilde{Y}_i$ and $\tilde{W}_i$ respectively. The calculation of the centroid method for X_{ic}, Y_{ic}, and W_{ic} to formulas that are:

$$X_{ic} = \frac{\int_{-\infty}^{\infty} x\mu_{\tilde{x}i}(x)dx}{\int_{-\infty}^{\infty} \mu_{\tilde{x}i}(x)dx} = \frac{1}{3}(X_{il} + X_{im} + X_{iu}) \quad (9)$$

$$Y_{ic} = \frac{\int_{-\infty}^{\infty} Y\mu_{\tilde{Y}i}(y)dy}{\int_{-\infty}^{\infty} \mu_{\tilde{Y}i}(y)dy} = \frac{1}{3}(Y_{il} + Y_{im} + Y_{iu}) \quad (10)$$

$$W_{ic} = \frac{\int_{-\infty}^{\infty} w\mu_{\tilde{w}i}(w)dw}{\int_{-\infty}^{\infty} \mu_{\tilde{w}i}(w)dw} = \frac{1}{3}(W_{il} + W_{im} + W_{iu}) \quad (11)$$

At this stage of the fuzzy environment, the α-cuts method is used for all the exogenous variables and error terms. As in the defuzzifation stage, the values of fuzzy are converted to output of the value of crisp as follows :

$$\overline{X}_{j_{sp_{i=1,2,3,4,5}}} = \sum_{i=1}^{j} \frac{(x_{j_{l_{sp}}}, x, x_{j_{u_{sp}}})}{j} \quad (12)$$

4 MONTE CARLO SIMULATION OF FMSPSSM

The purpose of the Monte Carlo simulation is used to calculate the values of the sample selection model Nawata (1994). The fuzzy model for Monte Carlo simulation as follows :

$$\tilde{y}_{i_{SP}} = \beta_0 + \beta_1 \tilde{x}_{i_{SP}} + \tilde{u}_{i_{SP}} \quad (13)$$

$$d_i = 1(\alpha_0 + \alpha_1 \tilde{w}_{i_{SP}} + \tilde{v}_{i_{SP}} \leq 0) \quad (14)$$

The following items are considered in the Monte Carlo study as follows :

- The effect of the correlation of $\tilde{x}_{i_{sp}}$ and $\tilde{w}_{i_{sp}}$
- The effect of the correlation of $\tilde{u}_{i_{sp}}$ and $\tilde{v}_{i_{sp}}$

$\tilde{x}_{i_{sp}}$ and $\tilde{w}_{i_{sp}}$ are the exogenous variables and the values are as follows :

$$\tilde{w}_{i_{sp}} = \tilde{s}_{1i_{sp}} , \quad (15)$$

$$\tilde{x}_{i_{sp}} = \frac{\lfloor \pi\tilde{s}_{1i_{sp}} + (1-\pi)\tilde{s}_{2i_{sp}} \rfloor}{\sqrt{\pi^2 + (1-\pi)^2}} \quad (16)$$

$\tilde{s}_{1i_{sp}}$ and $\tilde{s}_{2i_{sp}}$ are independent and identifically distributed (iid) random variables distributed uniformly on (0,20].

The values of the exogenous variables, $\tilde{x}_{i_{sp}}$ and $\tilde{w}_{i_{sp}}$ are as follows :

$\tilde{s}_{1i_{sp}}$ and $\tilde{s}_{2i_{sp}}$ are random uniform variables with mean 0 and variance 20. π is the correlation coefficient of $\tilde{x}_{i_{sp}}$ and $\tilde{w}_{i_{sp}}$ with π = 0.0 is considered. The fuzzy error terms $\{(\tilde{u}_{i_{sp}}, \tilde{v}_{i_{sp}})\}$ are jointly normal and determined as follows :

$$\tilde{v}_{i_{sp}} = \tilde{\varepsilon}_{1i_{sp}} \quad , \tag{17}$$

$$\tilde{u}_{i_{sp}} = \frac{\rho_0 \tilde{\varepsilon}_{1i_{sp}} + (1-\rho)\tilde{\varepsilon}_{2i_{sp}}}{\sqrt{\rho_0{}^2 + (1-\rho_0)^2}} \tag{18}$$

$\{\tilde{\varepsilon}_{1i_{sp}}\}$ are normal random variables with mean zero and variance 1. $\{\tilde{\varepsilon}_{2i_{sp}}\}$ are i.i.d. normal random variables with mean zero and variance 100. $\{\tilde{\varepsilon}_{1i_{sp}}\}$ and $\{\tilde{\varepsilon}_{2i_{sp}}\}$ are independently distributed. ρ_0 is the correlation coefficient of $\tilde{u}_{i_{sp}}$ and $\tilde{v}_{i_{sp}}$ with values of $\rho_0 = 0.0$ is considered. The error terms are calculated twice, which is for the classical error terms as well as the fuzzy error terms. ρ from [-0.99, 0.99] with interval 0.01. Our hypothesis is H_0: $\beta_1 = 0$ against H_1: $\beta_1 \neq 0$. If the hypothesis testing fail to reject $H_{0,}$ meaning that the model is not reflected our data. The sample sizes n=100, 200, and 400 are considered with 1000 replications on each sample size.

5 RESULT

The result is shown from Table 1 to Table 2 is the Mean, SD, RMSE with n=100, 200 and 400 for β_1**.** The first column shows the sample size (n), while the second column shows the parameter of bandwidth with a value of 0.2, 0.4, 0.6, 0.8, 1. Columns 3, 4, 5 provide information about the Mean, SD (standard deviation) and RMSE (root mean square error). This calculation is obtained by using DWADE estimator of the Powell estimator method.

Fuzzy Modified Semi-Parametric Sample Selection Model (FMPSSM)				
Sample Size (n)	**Bandwidth (h)**	**Mean**	**SD**	**RMSE**
100	0.2	1.071450	0.2689982	0.2783256
	0.4	1.070610	0.2614857	0.2708515
	0.6	1.070836	0.2563274	0.2659351
	0.8	1.071685	0.2526506	0.2626234
	1.0	1.072515	0.2499261	0.2602336
200	0.2	1.047969	0.1748553	0.1813158
	0.4	1.048289	0.1687668	0.1755394
	0.6	1.050418	0.1673359	0.1747662
	0.8	1.052373	0.1670478	0.1750656
	1.0	1.054159	0.1668233	0.1753945
400	0.2	1.016707	0.1283052	0.1293884
	0.4	1.015019	0.1254655	0.1263612
	0.6	1.011026	0.1228866	0.1238683
	0.8	1.015830	0.1212995	0.1223281
	1.0	1.016037	0.1200999	0.1211659

Table 1 : Fuzzy Modified Semi-Parametric Sample Selection Model (FMSPSSM) using π=0, ρ_0=0, and α-cut=0.2 for β_1

Fuzzy Modified Semi-Parametric Sample Selection Model (FMSPSSM)				
Sample Size (n)	**Bandwidth (h)**	**Mean**	**SD**	**RMSE**
100	0.2	1.060961	0.2578526	0.2649608
	0.4	1.057912	0.2520313	0.2585993
	0.6	1.056710	0.2469395	0.2533675
	0.8	1.057172	0.2448945	0.2514795
	1.0	1.057583	0.2436515	0.2503635
200	0.2	1.047969	0.1748553	0.1813158
	0.4	1.048289	0.1687668	0.1755394
	0.6	1.050418	0.1673359	0.1747662
	0.8	1.052373	0.1670478	0.1750656
	1.0	1.054159	0.1668233	0.1753945
400	0.2	1.016778	0.1278666	0.1289627
	0.4	1.015081	0.1250520	0.1259581
	0.6	1.015619	0.1225069	0.1234985
	0.8	1.015885	0.1209339	0.1219728
	1.0	1.016090	0.1197439	0.1208201

Table 2 : Fuzzy Modified Semi-Parametric Sample Selection Model (FMSPSSM) using π=0, ρ_0=0, and α-cut=0.4 for β_1

6 CONCLUSION

To reduce the problem of uncertainty that exists in MSPSSM, the fuzzy concept and its properties are the best alternative to overcome this problem. The triangular fuzzy number with membership function is used. The FMSPSSM has used the bandwidth by Powell estimator. Table 1 and Table 2 are calculation of the FMSPSSM for Mean, SD, RMSE under normality assumption. The values of SD and RMSE decreased with increased sample size, and also decreases with increasing values of α-cut.

References

[1] Andrews, D. W. K., *Asymptotic normality of series estimators for nonparametric and semiparametric regression models*. Econometrica, 59 No. 2, 307-345, 1991.

[2] Barnett, W. A., Powell, J. L. & Tauchen, G., *Nonparametric and semiparametric method in econometrics and statistics*. Cambridge University Press, 1991.

[3] Begun, J., Hall, W., Huang, W., & Wllner, J. *Information and asymptotic Efficiency in parametric-nonparametric models*. Annals of Statistics, 11, 432-452, 1983.

[4] Cosslett, S., *Semi-parametric estimation of a regression models with sample selectivity*. Cambridge University Press, 175-198, 1991.

[5] Honore, B. E., & Powell, J. L., *Pairwise difference estimators of censored and truncated regression models*. Journal of Econometrics, 64, 241-278, 1994.

[6] Nawata, K., *Estimation of sample selection bias models by maximum likelihood estimator and Heckman's two-step estimator*. Econometrics Letters, 45, 33-40, 1994.

[7] Powell, J. L., Stock, J. H., & Stoker, T. M., *Semi-parametric estimation of index coefficients*. Econometrica, 57, 1403-1430, 1989.
[8] Powell, J. L., *Semi-parametric estimation of bivariate latent variable models*. Social Systems Research Institute. University of Wisconsin-Madison, Working paper No.8704. 1987.
[9] Powell, J. L., *Semi-parametric estimation of censored selection models*. Unpublished manuscript. University of Wisconsin-Madison, 1989.
[10] Powell, J. L., *Estimation of semi-parametric models*. Handbook of Econometrics, Volume IV. Amsterdam., Elsevier Science Publishers, 1994.

[11] Safiih, M. L., Anton, A. B. K., & Abu Osman, M. T. , *Fuzzy semi-parametric sample selection model case study for participation of married women*. WSEAS Transactions on Mathematics, 7(3), 2008.
[12] Wu, C. O. , *The effects of kernel choices in density estimation with biased data*. Statistics & Probability Letters, 34, 373-383, 1997.
[13] Zadeh, L. A., *Fuzzy sets*. Information and Control, 8(3, 338-353), 1965.
[14] Zadeh, L. A., *The concept of a linguistic variable and its application to approximate reasoning – I*. Information Sciences, 8, 199-249, 1975.
[15] Zheng, J. X., *Semiparametric efficiency bounds for the binary choice and sample selection models under conditional symmetry*. Economics Letters, 47, 249-253, 1995.
[16] Zimmermann, H. J., *Fuzzy sets theory and its applications*, Kluwer academic Publishers, Second Edition, 1991.