Monte Carlo Simulation for Modified Parametric Of Sample Selection Models Through Fuzzy Approach

Yaya Sudarya Triana

Information Systems Department Information System, Universitas Mercu Buana Jakarta, Indonesia

ABSTRACT

The sample selection model is a combination of the regression and probit models. The models are usually estimated by Heckman's two-step estimator. However, Heckman's two-step estimator often performs poorly. In the context of parametric methods, the sample selection model is studied. The best approach is to take advantage of the tools provided by the theory of fuzzy sets. It appears very suitable for modeling vague concepts. It is difficult to determine some of the criteria and arrive at a quantitative value. Fuzzy sets theory and its properties through the concept of fuzzy number. The fuzzy function used for solving uncertain of a parametric sample selection model. Estimates from the fuzzy are used to calculate some of equation of the sample selection model. Finally, estimates of the Mean, Root Mean Square Error (RMSE) and the other estimators can be obtained by Heckman two-step estimator through iteration from some parameters and some of values.

Keywords: Fuzzy, Heckman's, Monte Carlo, Sample selection model, Simulation

*Correspondence to Author:

Yaya Sudarya Triana
Information Systems Department, Information System, Universitas Mercu Buana Jakarta, Indonesia

I. INTRODUCTION

The sample selection was developed by [1]. Sample selection

is widely used in various fields of economics. An early discussion of the problem of self-selectivity was that of [2], who discussed the problem of individuals selecting between hunting and fishing, based on their comparative advantage. The observed distribution of incomes of hunters and fishermen was defined by these choices [3].

The self-selectivity model that focused on selectivity bias was discussed by [4]. [5] introduced a two-step selection model, known as the Heckman two-step sample selection model. The sample selection bias problem in the context of decision by women to participate in the labour force is discussed by [4], [6], [1], [7], [8], [9] and [10].

II. PARAMETRIC SAMPLE SELECTION

MODEL [11] has proposed the parametric sample selection model as follows :

$$y_i^* = x_i'\beta + u_i$$
$$d_i^* = w_i'\alpha + v_i \qquad (1)$$
$$y_i = y_i^* d_i$$

$$d_i = \begin{cases} 1 & \text{if} & d_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

The variables y_i^* and d_i^* are unobserved,

whereas yi is observed. y_i^* and d_i^* are dependent variables, xi and wi are vectors of independent variables, α and β are unknown parameter vectors, ui and vi are error terms.

In Equation (1), there are error terms (u, v) which are usually correlated, so that the regression of y on x will not a give consistent estimates of $\beta 0$ and $\beta 1$. The approach of the error terms (ui, vi) are assumed to follow a bivariate normal distribution. It is commonly assumed that ui and

vi have a bivariate normal distribution :

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim BN \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix} \right]$$

According to [12] and [13], there are two parts in Equation (1). The first part is participation equation (a binary decision equation). The second part is the wage equation (outcome equation or selection part). Independent variable xi usually contain at least one variable which does not appear in variable wi. The outcome equation describes the relationship between the dependent variable yi and independent variable xi, whereas in the selection equation describes the relationship between the dependent variable di and the independent variable wi.

III. ALPHA-CUT OF FUZZY NUMBER

Alpha cuts are simply threshold levels that convert a fuzzy set

into a crisp set. The process of converting a fuzzy set to a crisp one is called defuzzification. An alpha-cut A of a fuzzy number A is defined as the set $\{x \in R \mid A(x) \geq \alpha\}$. A is completely determined by the collection $(A_\alpha)\alpha \in [0,1]$. An alpha cut is the behaviour sensitivity of the system to the behavior under observation. At some point, as the information value diminishes, one no longer wants to be "bothered" by the data. In many systems, due to the inherent limitations of the mechanisms of observation, the information becomes suspect below a certain level of reliability.

The membership function is cut horizontally at a finite number of α-levels between 0 and 1. For each α-level of the parameter, the model is run to determine the minimum and maximum possible values of the output. This information is then directly used to construct the corresponding fuzziness (membership function) of the output which is used as a measure of uncertainty. If the output is monotonic with respect to the dependent fuzzy variable, the process is rather

simple since only two simulations will be enough for each α-level (one for each boundary). Otherwise, optimization routines have to be carried out to determine the minimum and maximum values of the output for each α-level. Figure 1 shows an illustration of the alpha-cut of triangular fuzzy number.
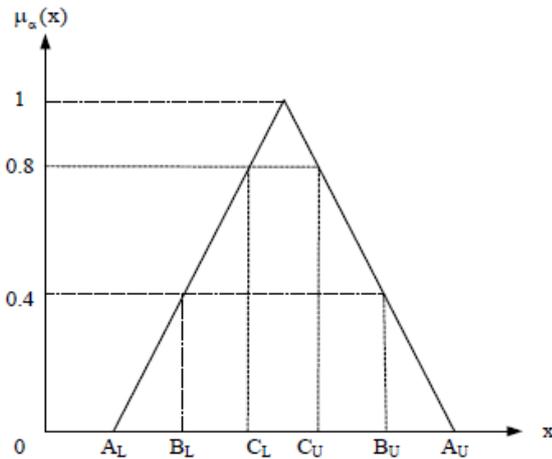

Figure 1 The alpha cut of triangular fuzzy number

From Figure 1, the confidence fuzzy interval defined by different value of alpha cut. For example : α(0.4) and (0.8), then their confidence fuzzy interval are [BL, BU] and [CL, CU]. This relationship denoted by (α(0.4), [BL, BU]), and (α(0.8), [CL, CU]) with [BL, BU] ≥ [CL, CU].

IV. THE MODEL

In this section the model of participant, non-participant, and combination of participant and non-participant are discussed. The model is written as follows :

$$y_i = \begin{cases} y_0, & \text{if } d_i = 0 \\ y_1, & \text{if } d_i = 1 \end{cases}$$

$$d_i = \begin{cases} 1, & \text{if } w_i\alpha + v_i > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where di is a selection equation of the first stage. The values of di are 0 and 1 which is di = 1 for participant and di = 0 for nonparticipant. yi is dependent variable of the outcome equation. In the Equation (2) is derived into two, ie y0 and y1. y0 is refered to the outcome of the Equation (2) is non-participant, whereas y1 for participant. xi is independent variable. The details about the

equation of y1 and y0 are as follows :

$$y_{0i}^* = x_{0i}'\beta_0 + u_{0i} \quad \text{if} \quad d_i = 0 \quad (3)$$

$$y_{1i}^* = x_{1i}'\beta_1 + u_{1i} \quad \text{if} \quad d_i = 1 \quad (4)$$

where u0i and u1i are error terms. The outcome equation for non-participant and participant in Equation (2), can be summarized as follows :

$$Y_{i0} = y_{0i}^*(1 - d_i) \quad \text{for non-participant} \quad (5)$$

$$y_{i1} = y_{1i}^* d_i \quad \text{for participant} \quad (6)$$

hence, the combination of both non-participant and participant, will generate the equation, as follows :

$$y_i = y_{0i}^*(1 - d_i) + y_{1i}^* d_i \quad (7)$$

V. MONTE CARLO SIMULATION OF PARAMETRIC SAMPLE SELECTION MODEL

The purpose of the Monte Carlo simulation is used to calculate the values of the sample selection model [14]. The fuzzy model for Monte Carlo simulation is as follows :

$$\tilde{y}_i = \beta_0 + \beta_1 \tilde{x}_i + \tilde{u}_i \quad (8)$$

$$d_i = 1(\alpha_0 + \alpha_1 \tilde{w}_i + \tilde{v}_i \leq 0) \quad (9)$$

The following items are considered in the Monte Carlo study: - The effect of the correlation of $\tilde{x}_i$ and $\tilde{w}_i$

- The effect of the correlation of $\tilde{u}_i$ and $\tilde{v}_i$

$$\tilde{W}_i = \tilde{s}_{1i} , \quad (10)$$

$$\tilde{x}_i = \frac{[\pi \tilde{s}_{1i} + (1 - \pi)\tilde{s}_{2i}]}{\sqrt{\pi^2 + (1 - \pi)^2}} \quad (11)$$

$\tilde{s}_{1i}$ and $\tilde{s}_{2i}$ are independent and identically distributed (i.i.d.) random variables distributed uniformly on (0,20]. The values of the exogenous variables, $\tilde{x}_i$ and $\tilde{w}_i$ are as follows :

$\tilde{s}_{1i}$ and $\tilde{s}_{2i}$ are random uniform variables with mean = 0 and variance = 20. π is the correlation coefficient of $\tilde{x}_i$ and $\tilde{w}_i$ with π = 0.0, 0.5, and

0.9 are considered. The fuzzy error terms{($\tilde{u}_i$,$\tilde{v}_i$)} are jointly normal and determined as follows:

$$\widetilde{v}_i = \widetilde{\varepsilon}_{1i} \; , \qquad (12)$$

$$\widetilde{u}_i = \frac{\rho_0 \widetilde{\varepsilon}_{1i} + (1-\rho)\widetilde{\varepsilon}_{2i}}{\sqrt{\rho_0^2 + (1-\rho_0)^2}} \qquad (13)$$

{$\tilde{\varepsilon}_{1i}$ }are normal random variables with mean = 0 and variance = 1. {$\tilde{\varepsilon}_{2i}$} are i.i.d. normal random variables with mean = 0 and variance = 100. {$\tilde{\varepsilon}_{1i}$ } and {$\tilde{\varepsilon}_{2i}$} are independently distributed. ρ_0 is the correlation coefficient of $\tilde{u}_i$ and $\tilde{v}_i$ with values of ρ_0 = 0.0, 0.5, and 0.9 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. The true values of the parameters are β0 = -10.0, β1 = 1.0, α0 = -1.0, and α1 = 0.1,

Our hypothesis is H0: β1 = 0 against H1: β1 ≠ 0. If the hypothesis testing fails to reject H0, meaning that the model does not reflect our data. The sample sizes n=100, 200, and 400 are considered with 1000 replications on each sample size.

True value is the actual variation that would be measured. In this case, so that the expected results are similar to the entered value, the true value for β0 and β1 are   -10.0 and 1.0 is expected to result from these estimates and values. The chosen sample size, n = 100, 200 and 400 are only example values, this value is expected to meet the minimum value that should be taken, should we take another value, e.g. 101, 213, 379, etc. The larger the sample size and replication, the better. So as the normal random value is generated, this value is exemplified by a normal distribution (0.1) which means normal standards of raw materials, normal distribution (0.100) and the uniform normal distribution (0.20). This simulation will be measured in the range and value, so hopefully the results will be in the range of the accepted values.

VI. RESULT

The calculation of Monte Carlo simulation from Table 1 to

Table 5 using fuzzy α-cut 0.2, 0.4, 0.5, 0.6, 0.8, and 1.0. The effect of correlation can be viewed from the fuzzy exogenous variables between $\tilde{w}$ and $\tilde{x}$ and the effect of correlation of fuzzy error terms between $\tilde{u}$ and $\tilde{v}$, and the effects of comparison among several different sample sizes are n = 100, n = 200, and n = 400, shown in the table on the columns 1, 2 and 3, while columns 4, 5, 6 and 7, 8, 9 show Mean, SD and RMSE of parameter β0 and β1, where SD is standard deviation and RMSE is root mean square errors.

Result is shown from Table 1 to Table 5 are the Mean, SD, RMSE with n=100, 200 and 400 for parameters of β0, β1. From this table, represented the α-cuts of triangular fuzzy number with a value of 0.2, 0.4, 0.5, 0.6, and 0.8. The first column shows the parameter of phi with values 0.0, 0.5, 0.9, while the second column shows the parameter of rho with a value of 0.0, 0.5, 0.9.

Table 1 display that the mean of parameters of β0 and β1 showed about the consistency information of the parameter estimator that approaches the true values, while the SD and RMSE provide information about the level of efficiency for the parameters in the estimation. Sample size n = 100, n = 200 and n = 400 provides the information that the larger the sample size, the smaller the values of SD and RMSE, it means that more efficient and accurate the estimator, if there is no relationship between fuzzy error term.

For example in Table 1, for the sample size n = 100, the values of mean of parameters β0 = -10.041 and β1 = 1.004, the values of SD = 4.325, RMSE = 4.325 of parameter of β0 and the values of SD = 0.260, RMSE = 0.260 of parameter of β1, for the sample size n = 200, the values of mean of parameters   β0 = -9.831 and β1 = 0.991,

the values of SD = 2.964, RMSE = 2.969 of parameter of β0 and the values of SD = 0.171, RMSE = 0.171 of parameter of β1, while for the sample size n = 400, the values of mean of parameters β0 = -9.969 and β1 = 1.001, the values of SD = 2.116, RMSE = 2.116 of parameter of β0 and the values of SD = 0.121, RMSE = 0.121 of parameter of β1.

Result of Mean, SD, RMSE of FMPSSM with n=400 for β0, β1 and α-cut = 0.2 From Table 1, it can be shown that for the α-cut = 0.2, with sample size n = 100, where values of ρ0 and π are 0 (meaning that there is no relationship between ρ and π), so the value becomes small of SD and RMSE for parameter of β0, the values of SD = 4.325, RMSE = 4.325 and for parameter of β1, the values of SD = 0.260, RMSE = 0.260. When the sample size increases and become n = 200, SD and RMSE values for the parameters β0 and β1 decreases, i.e. for value of SD = 2.964, RMSE = 2.969 of parameter β0, and the value of SD = 0.171, RMSE = 0.171 of the parameter β1. When the sample size increases again for n = 400, SD and RMSE values for the parameters of

β0 and β1 is smaller again, that is the values of SD = 2.116, RMSE = 2.116 for parameter of β0 and the values of SD = 0.121, RMSE = 0.121 for parameter of β1. While ρ0 increases to moderate values of 0.5 (moderate correlation between fuzzy error term), then the values of SD and RMSE becomes larger than when the value of rho = 0 (no error relationship between fuzzy terms).

Increase in value of ρ0 as shown in Table 1 indicates that the greater correlation of fuzzy error terms between $\tilde{u}$ and $\tilde{v}$, the smaller the value of SD and RMSE. From the above table also illustrates, that when the value of ρ0 and π strong (ρ0 and π = 0.9), then the values of SD and RMSE increased, compared with the value of which no correlation of ρ0 and π are moderate. This shows that the occurrence of multicollinearity of fuzzy exogenous variables between

$\tilde{w}_I$ and $\tilde{x}_i$, but the value of this multicollinearity will be corrected by sample size that continues to expand.

Table 1 : Mean, SD, RMSE of FMPSSM under normality assumption with n=100, 200, 400 for β0, β1 and α-cut = 0.2

π	ρ0	n	β0			β1		
			Mean	SD	RMSE	Mean	SD	RMSE
0.0	0.0	100	-10.041	4.325	4.325	1.0039	0.2598	0.260
		200	-9.831	2.964	2.969	0.991	0.171	0.171
		400	-9.969	2.116	2.116	1.001	0.121	0.121
	0.5	100	-9.759	6.628	6.633	0.999	0.358	0.358
		200	-9.966	4.594	4.594	1.011	0.265	0.265
		400	-9.967	3.133	3.133	0.999	0.178	0.178
	0.9	100	-9.773	6.244	6.249	1.009	0.323	0.324
		200	-9.861	4.074	4.077	1.001	0.224	0.224
		400	-9.967	2.835	2.835	1.007	0.159	0.159
0.5	0.0	100	-9.967	4.151	4.151	1.017	0.364	0.364
		200	-9.959	2.716	2.716	1.001	0.245	0.245
		400	-10.079	1.876	1.877	1.015	0.171	0.172
	0.5	100	-9.604	6.036	6.049	1.002	0.532	0.532
		200	-9.876	4.023	4.025	1.006	0.371	0.371
		400	-9.960	2.911	2.911	1.005	0.253	0.253
	0.9	100	-9.690	5.555	5.563	1.006	0.437	0.437
		200	-9.785	3.760	3.766	1.011	0.327	0.327
		400	-10.009	2.553	2.553	0.997	0.227	0.227
0.9	0.0	100	-10.038	11.772	11.772	1.012	1.799	1.799
		200	-10.149	5.665	5.667	0.996	1.239	1.239
		400	-9.999	3.345	3.345	1.010	0.880	0.880
	0.5	100	-9.648	14.229	14.233	1.024	2.522	2.522
		200	-9.555	7.524	7.537	1.057	1.816	1.817
		400	-9.542	4.800	4.821	1.101	1.292	1.296
	0.9	100	-10.032	13.081	13.081	0.900	2.364	2.366
		200	-10.046	6.554	6.554	0.957	1.505	1.506
		400	-9.705	4.165	4.175	0.975	1.093	1.093

Table 1 display that the mean of parameters of β0 and β1 showed about the consistency information of the parameter estimator that approaches the true values, while the SD and RMSE provide information about the level of efficiency for the parameters in the estimation. Sample size n = 100, n = 200 and n = 400 provides the information that the larger the sample size, the smaller the values of SD and RMSE, it means that more efficient and accurate the estimator, if there is no relationship between fuzzy error term.

For example in Table 6.1, for the sample size n = 100, the values of mean of parameters β0 = -10.0969 and β1 = 0.9967, the values of SD = 4.6709, RMSE = 4.6719 of parameter of β0 and the values of SD = 0.2530, RMSE = 0.2530 of parameter of β1, for the sample size n = 200, the values of mean of parameters β0 = -9.9669 and β1 = 0.9984, the values of SD = 3.0880, RMSE = 3.0882 of parameter of β0 and the values of SD = 0.1737, RMSE = 0.1737 of parameter of β1, while for the sample size n = 400, the values of mean of parameters β0 = -9.9701 and β1 = 1.0020, the values of SD = 2.1179, RMSE = 2.1182 of parameter of β0 and the values of SD = 0.1249, RMSE = 0.1249 of parameter of β1.

From Table 1, it can be shown that for the α-cut = 0.2, with sample size n = 100, where values of ρ0 and π are 0 (meaning that there is no relationship between ρ and π), so the value becomes small of SD and RMSE for parameter of β0, the values of SD = 4.6709, RMSE = 4.6719 and for parameter of β1, the values of SD = 0.2530, RMSE = 0.2530. When the sample size increases and become n = 200, SD and RMSE values for the parameters β0 and β1 decreases, i.e. for value of SD = 3.0880, RMSE = 3.0882 of parameter β0, and the value of SD = 0.1737, RMSE = 0.1737 of the parameter β1. When the sample size increases again for n = 400, SD and RMSE values for the parameters of β0 and β1 is smaller again, that is the values of SD = 2.1179, RMSE = 2.1182 for parameter of β0 and the values of SD = 0.1249, RMSE = 0.1249 for parameter of β1. While ρ0 increases to moderate values of 0.5 (moderate correlation between fuzzy error term), then the values of SD and RMSE becomes larger than when the value of rho = 0 (no error relationship between fuzzy terms).

Table 2 : Mean, SD, RMSE of FMPSSM under normality assumption with n=100, 200, 400 for β0, β1 and α-cut = 0.4

π	ρ0	n	β0			β1		
			Mean	SD	RMSE	Mean	SD	RMSE
0.0	0.0	100	-9.775	4.539	4.545	0.988	0.249	0.249
		200	-9.911	3.191	3.193	0.999	0.180	0.180
		400	-10.085	2.071	2.073	1.003	0.121	0.121
	0.5	100	-9.515	6.681	6.698	0.982	0.386	0.386
		200	-10.077	4.530	4.530	1.006	0.259	0.259
		400	-10.076	3.225	3.226	0.997	0.180	0.180
	0.9	100	-9.978	5.893	5.893	1.005	0.322	0.322
		200	-9.630	4.182	4.199	0.991	0.230	0.231
		400	-9.972	2.865	2.865	1.003	0.157	0.157
0.5	0.0	100	-10.108	4.383	4.385	1.000	0.363	0.363
		200	-10.065	2.685	2.686	1.006	0.256	0.256
		400	-9.924	1.787	1.789	0.995	0.174	0.174
	0.5	100	-10.071	6.140	6.140	1.001	0.539	0.539
		200	-9.872	4.042	4.044	0.998	0.365	0.365
		400	-10.014	2.850	2.851	0.997	0.252	0.252
	0.9	100	-9.730	5.379	5.386	0.995	0.449	0.449
		200	-10.037	3.785	3.785	0.990	0.311	0.311
		400	-10.065	2.513	2.514	1.008	0.222	0.222
0.9	0.0	100	-10.267	10.585	10.589	1.011	1.855	1.855
		200	-9.653	5.499	5.509	1.066	1.217	1.219
		400	-10.024	3.381	3.381	0.978	0.884	0.884
	0.5	100	-9.198	13.551	13.575	0.982	2.595	2.595
		200	-9.931	9.042	9.043	0.945	1.868	1.869
		400	-10.145	4.712	4.714	0.946	1.278	1.279
	0.9	100	-9.866	11.909	11.910	1.036	2.252	2.252
		200	-9.313	7.034	7.067	1.121	1.593	1.598
		400	-10.096	4.100	4.101	0.946	1.109	1.110

Table 3 : Mean, SD, RMSE of FMPSSM under normality assumption with n=100, 200, 400 for β0, β1 and α-cut = 0.5

π	ρ_0	n	β_0			β_1		
			Mean	SD	RMSE	Mean	SD	RMSE
0.0	0.0	100	-10.047	4.555	4.555	1.007	0.261	0.261
		200	-10.061	3.002	3.002	0.998	0.183	0.183
		400	-9.999	2.131	2.131	1.004	0.123	0.123
	0.5	100	-9.831	6.682	6.684	1.011	0.361	0.361
		200	-9.889	4.497	4.498	0.998	0.261	0.261
		400	-10.053	3.149	3.150	1.008	0.179	0.179
	0.9	100	-9.454	6.387	6.410	0.995	0.318	0.318
		200	-9.907	3.991	3.992	0.993	0.218	0.218
		400	-9.962	2.858	2.858	1.003	0.153	0.153
0.5	0.0	100	-9.982	4.232	4.232	1.015	0.363	0.363
		200	-9.938	2.698	2.698	0.991	0.234	0.234
		400	-10.001	1.897	1.897	0.998	0.171	0.171
	0.5	100	-9.988	6.151	6.151	1.018	0.534	0.534
		200	-10.141	4.140	4.143	0.988	0.371	0.371
		400	-9.933	2.917	2.918	0.999	0.248	0.248
	0.9	100	-9.477	5.786	5.810	0.981	0.457	0.458
		200	-9.953	3.747	3.747	0.998	0.298	0.298
		400	-9.925	2.524	2.525	0.989	0.227	0.227
0.9	0.0	100	-10.271	9.747	9.751	0.967	1.779	1.780
		200	-9.993	5.421	5.421	1.001	1.269	1.269
		400	-10.087	3.095	3.096	0.969	0.862	0.862
	0.5	100	-9.104	16.068	16.093	1.069	2.627	2.627
		200	-9.780	7.767	7.770	1.033	1.873	1.873
		400	-9.900	4.685	4.687	1.014	1.244	1.244
	0.9	100	-10.217	12.761	12.763	0.889	2.231	2.234
		200	-9.944	6.868	6.869	1.041	1.554	1.554
		400	-9.855	4.267	4.270	0.997	1.102	1.102

Table 4 : Mean, SD, RMSE of FMPSSM under normality assumption with n=100, 200, 400 for β0, β1 and α-cut = 0.6

π	ρ_0	n	β_0			β_1		
			Mean	SD	RMSE	Mean	SD	RMSE
0.0	0.0	100	-9.961	4.339	4.340	1.000	0.253	0.253
		200	-9.960	3.148	3.148	1.010	0.180	0.181
		400	-9.922	2.042	2.043	1.000	0.123	0.123
	0.5	100	-9.944	6.614	6.614	1.003	0.369	0.369
		200	-10.224	4.475	4.481	1.012	0.254	0.255
		400	-1.000	3.051	3.051	0.997	0.175	0.175
	0.9	100	-9.782	6.230	6.233	1.001	0.328	0.328
		200	-9.998	3.980	3.980	0.996	0.222	0.222
		400	-9.915	2.735	2.736	0.994	0.161	0.161
0.5	0.0	100	-10.112	3.875	3.876	1.005	0.362	0.362
		200	-10.066	2.675	2.676	0.985	0.251	0.251
		400	-9.901	1.881	1.884	0.993	0.178	0.178
	0.5	100	-9.858	5.986	5.987	0.996	0.556	0.556
		200	-9.816	4.018	4.022	1.000	0.358	0.358
		400	-10.060	2.735	2.736	0.998	0.260	0.260
	0.9	100	-9.599	5.601	5.615	1.011	0.440	0.440
		200	-9.848	3.804	3.807	0.996	0.298	0.298
		400	-9.888	2.564	2.566	0.989	0.217	0.217
0.9	0.0	100	-10.226	11.238	11.241	0.959	1.761	1.761
		200	-10.105	5.645	5.646	0.967	1.293	1.293
		400	-10.091	3.203	3.205	1.023	0.851	0.851
	0.5	100	-10.206	13.940	13.942	0.923	2.572	2.573
		200	-9.901	8.333	8.333	1.026	1.829	1.829
		400	-10.052	5.030	5.030	0.902	1.331	1.335
	0.9	100	-10.008	13.863	13.863	0.946	2.307	2.307
		200	-10.068	6.653	6.653	0.929	1.581	1.583
		400	-9.852	4.309	4.311	1.030	1.142	1.142

Table 5 : Mean, SD, RMSE of FMPSSM under normality assumption with n=100, 200, 400 for β0, β1 and α-cut = 0.8

π	ρ0	n	β0			β1		
			Mean	SD	RMSE	Mean	SD	RMSE
0.0	0.0	100	-10.005	4.430	4.430	0.995	0.254	0.254
		200	-10.063	3.007	3.008	0.998	0.180	0.180
		400	-9.995	2.099	2.099	1.004	0.130	0.130
	0.5	100	-9.662	6.554	6.563	0.987	0.366	0.366
		200	-9.887	4.490	4.491	0.999	0.266	0.266
		400	-9.815	3.134	3.139	0.999	0.181	0.181
	0.9	100	-9.869	5.954	5.956	1.018	0.326	0.327
		200	-9.884	4.202	4.204	1.003	0.222	0.222
		400	-10.172	2.719	2.725	1.006	0.153	0.153
0.5	0.0	100	-9.886	3.999	4.000	0.996	0.356	0.356
		200	-10.121	2.758	2.761	0.991	0.245	0.246
		400	-9.969	1.883	1.884	1.007	0.176	0.176
	0.5	100	-9.803	6.135	6.138	0.996	0.523	0.523
		200	-10.068	4.029	4.030	0.996	0.369	0.369
		400	-10.062	2.766	2.766	0.988	0.251	0.251
	0.9	100	-9.766	5.583	5.588	1.022	0.451	0.452
		200	-9.795	3.692	3.698	0.985	0.316	0.316
		400	-9.902	2.486	2.488	0.998	0.214	0.214
0.9	0.0	100	-10.102	10.428	10.429	0.995	1.805	1.805
		200	-9.960	5.367	5.367	1.039	1.280	1.281
		400	-10.048	3.378	3.378	0.984	0.916	0.916
	0.5	100	-9.112	14.978	15.005	1.017	2.707	2.708
		200	-9.376	7.830	7.855	1.076	1.823	1.825
		400	-10.149	5.046	5.048	0.926	1.301	1.303
	0.9	100	-9.177	14.470	14.493	1.058	2.217	2.217
		200	-9.660	6.942	6.950	1.012	1.631	1.631
		400	-9.832	4.101	4.104	1.015	1.093	1.093

Table 5 : Mean, SD, RMSE of FMPSSM under normality assumption with n=100, 200, 400 for β0, β1 and α-cut =1.0

π	ρ0	n	β0			β1		
			Mean	SD	RMSE	Mean	SD	RMSE
0.0	0.0	100	-9.852	4.569	4.571	0.992	0.251	0.251
		200	-9.912	3.028	3.029	0.997	0.177	0.177
		400	-9.955	2.135	2.135	0.997	0.118	0.118
	0.5	100	-9.568	6.645	6.659	1.000	0.363	0.363
		200	-9.950	4.304	4.304	0.998	0.262	0.262
		400	-9.901	3.232	3.234	1.003	0.175	0.175
	0.9	100	-9.702	5.964	5.971	0.987	0.322	0.322
		200	-9.659	4.158	4.172	0.994	0.222	0.223
		400	-9.816	2.786	2.792	0.988	0.157	0.157
0.5	0.0	100	-9.947	4.060	4.060	0.985	0.345	0.345
		200	-10.030	2.733	2.733	1.000	0.249	0.249
		400	-10.028	1.909	1.909	0.998	0.169	0.169
	0.5	100	-9.620	5.933	5.945	0.996	0.509	0.509
		200	-9.788	4.135	4.140	1.000	0.369	0.369
		400	-10.122	2.887	2.889	1.006	0.259	0.259
	0.9	100	-9.752	5.728	5.733	1.008	0.446	0.446
		200	-9.668	3.823	3.837	1.002	0.319	0.319
		400	-9.999	2.543	2.543	1.001	0.213	0.213
0.9	0.0	100	-9.587	12.504	12.511	1.059	1.801	1.802
		200	-9.789	5.473	5.477	1.022	1.228	1.228
		400	-10.089	3.196	3.197	0.957	0.878	0.879
	0.5	100	-9.487	14.266	14.276	1.039	2.674	2.674
		200	-9.328	8.536	8.562	1.135	1.827	1.832
		400	-9.930	4.746	4.747	0.976	1.264	1.264
	0.9	100	-9.873	14.167	14.167	0.862	2.575	2.578
		200	-9.872	6.410	6.411	1.018	1.456	1.456
		400	-10.100	3.836	3.837	0.956	1.091	1.092

The consistency for FMPSSM and FMSPSSM has been discussed. The effect of the correlation of variables

$\tilde{x}_i$ and $\tilde{w}_i$, then the effect of the correlation of error terms $\tilde{u}_i$ and $\tilde{v}_i$

are then observed. The FMSPSSM has used the bandwidth by the Powell estimator. The effects of bandwidth changes are studied and researched, and later observed whether it is consistent or not.

Table 6.1 to Table 6.5 are calculation of FMPSSM for Mean, SD, RMSE under normality assumption, Table 6.6 to Table 6.10 are calculation of FMPSSM for Mean, SD, RMSE under non-normality assumption. The values of SD and RMSE

decreased with increased sample size, and also decreases with increasing values of α-cut. The values of SDn100 > SDn200 > SDn400 and the values of RMSEn100 > RMSEn200 > RMSEn400. Table 6.11 to Table 6.15 are the calculation of FMSPSSM for Mean, SD, RMSE. From the tables it appears that the values

of SD and RMSE decreased with an increasing sample size, the α-cut, and the bandwidth.

The results obtained from Monte Carlo simulations for the previous study, suggesting that FPSSM is consistent and efficient, the same as the results obtained with FMPSSM.

V. CONCLUSION

To reduce the problem of uncertainty that exists in the

parametric sample selection models, then created a fuzzy approach. fuzzy concept that used for modified of the parametric sample selection models provides an alternative for the handle to the problem when the model involves the characteristic vagueness, uncertainty and ambiguity.

Result of the Parametric Sample Selection Models using a fuzzy approach shows that the larger the sample size, the smaller the values of SD and RMSE, it means that the more efficient

and accurate the estimator, then the greater the value of α-cut, the smaller the values of SD and RMSE, and the more efficient and accurate.

REFERENCES

1. Heckman, J.J., "Shadow price, market wages and labor supply", Econometrics, 42, p. 679-694, 1974.
2. Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. Oxford Economic Papers, 3. 135-146.
3. Maddala, G.S., "Limited-dependent and qualitative in econometrics", Cambridge University Press. p. 257-289, 1983.
4. Gronau, R., "Wage comparisons: A selectivity bias", Journal of Political Economy, 82, p. 1119-1143, 1974.
5. Heckman, J.J., "Sample selection as a specification error", Econometrica, Vol.47, p.153-161, 1979.
6. Lewis, H.G., "Comments on selectivity biases in wage comparisons", Journal of Political Economy, 82, p. 1145-1155, 1974.
7. Neumark, D. (1988). Employers' Discriminatory Behavior and the Estimation of Wage Discrimination. Journal of Human Resource, 23,279-295.
8. Gerfin, M. (1996). Parametric and semi-parametric estimation of the binary response model of labour market participant. Journal of Applied Econometrics, 11, 321-339.
9. Vella, F., "Estimating models with sample selection bias: A survey", Journal of Human Resource, Vol. 33, p. 127-169, 1998.
10. Christofides, L. N., Li, Q., Liu, Z., & Min, I. (2003). Recent two-stage sample selection procedure with an application to the gander wage gap. Journal of Business & Economic Statistics, 21 (3), 396-405.
11. Heckman, J.J., "The common structure of statistical models of truncation, sample selection, and limited dependent variables, and a simple estimation for such models.", Annals of Economic and Social Measurement, 5, 475-492.
12. Schafgans, M. (1996). Semi-parametric estimation of a sample selection model: Estimation of the intercept, theory and applications.Unpublished Ph.D. Thesis. Yale University. New Haven. USA.
13. Martins, M. F. O. (2001). Parametric and semiparametric estimation of sample selection models: An empirical application to the female labour force in Portugal. Journal of Applied Econometrics, 16, 23-39.
14. Nawata, K., "Estimation of Sample Selection Bias Models by Maximum Likelihood Estimator and Heckman's Two-Step estimator", Econometrics Letters, 45, 33-40, 1994.